

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:01 ; Search time 8358.86 Seconds
(without alignments)
12307.996 Million cell updates/sec

Title: US-10-014-717-1_COPY_11872_16104

Perfect score: 4233

Sequence: 1 atgacgacatcagctctc.....agggcaggagacgtacgctaa 4233

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estcbsa.*
2: em_estchum.*
3: em_estcnu.*
4: em_estmu.*
5: em_estrov.*
6: em_estropl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_dln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rtd.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	ID	Description
1	175.4	4.1	812	29	B2671349 PUBKS59TD
2	174.4	4.1	627	29	B2563956 pac62-164
3	157.6	3.7	1552	29	B2572480 msh2_2654
4	133	3.1	765	29	B2562842 pac62-164

C	5	128.8	3.0	646	28	A0989512
	6	125	3.0	1117	29	B2548527
	7	121.6	2.9	921	29	B2570019
	8	120.8	2.9	1038	29	B2567332
	9	119	2.8	667	29	B2548604
	10	117	2.8	753	29	B2561650
	11	114.4	2.7	840	29	B2558891
	12	114	2.7	827	29	B2680689
	13	111.8	2.6	901	29	B2574671
	14	110.4	2.6	889	29	B2551143
	15	109.2	2.6	633	29	B2551440
	16	107.8	2.5	827	29	B2559631
	17	107.6	2.5	648	29	B2548537
	18	107	2.5	709	29	BH488170
	19	104	2.5	1429	28	B2575585
	20	101.6	2.4	688	29	B2567972
	21	98.6	2.3	792	29	B2574700
	22	97	2.3	780	29	B2573398
	23	96.2	2.3	1342	29	B2556627
	24	96	2.3	1122	29	B2558365
	25	94.2	2.2	778	29	B2573262
	26	93.4	2.2	915	29	B2568304
	27	93.2	2.2	1307	29	B2576975
	28	92.8	2.2	628	12	B1180727
	29	92.2	2.2	580	29	B2561162
	30	91.6	2.2	885	29	B2551313
	31	91.6	2.2	1409	29	B2577581
	32	90.6	2.1	610	29	P974L
	33	90.2	2.1	1388	29	B2555697
	34	89.6	2.1	1118	29	B2551441
	35	89.6	2.1	1150	29	B2578915
	36	89	2.1	891	29	B2567106
	37	88.8	2.1	850	29	B2670377
	38	87.8	2.1	1131	29	B2559111
	39	87	2.1	1489	29	B2572227
	40	85.6	2.0	1154	29	B2553289
	41	84.8	2.0	1364	29	B2567613
	42	84.2	2.0	1594	29	B2551753
	43	83.8	2.0	715	29	B2560475
	44	83.8	2.0	807	29	B2568590
	45	82.8	2.0	881	29	B2549419

ALIGNMENTS

RESULT 1
LOCUS B2671349 812 bp DNA linear GSS 05-FEB-2003
DEFINITION PUBKS59TD ZM 0.6 1.0 KB zea mays genomic clone ZMWB78077021,
genomic survey sequence.

ACCESSION B2671349
VERSION B2671349.1 GI:28219323
KEYWORDS GSS.

SOURCE ORGANISM

Zeae mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS White, C.A., Quackenbush, J., Van Aken, S., Uterback, T., Resnick
'A', Frazer, C.M., Yuan, Y., San Miguel, P., Ma, D. and Bennettzen, J.
Maize Genomics Consortium
Unpublished
Contact: Cathy White, Tel: 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitec@ncicr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers

FEATURES

Db	536	GGTGTGGCCGGACAGCTCGCGCTTTACCTCGGGCGAGCCACGAGCGCGCATCTGGC	595
Oy	2477	CCATCGGGTACCC	2489
Db	596	CGATCGAACATCC	608
RESULT 3			
B2572480/c		1552 bp	DNA
LOCUS			linear
DEFINITION			SS 17-DEC-2002
ACCSSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
1..1552			
/organism="Pseudomonas aeruginosa"			
/mol_type="genomic DNA"			
/strain="MSH"			
/db_xref="taxon:287"			
/clone="msh2_2654"			
/note="Environmental isolate. Whole genomic shotgun library."			
BASE COUNT			
320 a			
448 c			
387 g			
390 t			
7 others			
Query Match			
3.7%			
Score 157.6;			
DB 29;			
Length 1552;			
Best Local Similarity 54.4%;			
Prod. No. 1.3e-22;			
Matches 356;			
Conservative 0;			
Mismatches 275;			
Indels 24;			
Gaps 1;			
Oy	1653	CGAAGAGCTTTCGCGCGCTTTCGCGCGCGCATTTGGCGCGCGCGCTGGCGGAGCAGGGGGCAGC	1712
Db	724	CGAGGAACTGTCGAGCGTGCATCTGCGCATGCGCGCGCTTTCGCGCGGAGCGGGGTGCG	665
Oy	1713	CCCGAACAATTTGTCGCGGTGTGTATGAGAAAGCTGGAGCAGGTGTCGCGGTCT	1772
Db	664	CCCTGGCGACGCGGTGCGAGTCACTGCTCGCGCGCGAGCCGACGACGAGGTGCGGCGGTATT	605
Oy	1773	CGCGGTCTCGATGACGCGCGGCGCTTACTGTCGATGCAATGCGGACCTTACCGGCGGAGCG	1832
Db	604	CGGGGTGTCGCGCGGACGCGCTGCTTAAGTGTGCTGACATGACCAAGCCGCGCGCAG	545
Oy	1833	TATCCACTACCTCCCTCGATCATGTGAGTAAAGCTCGTGAGAGCGACCATGCTGGA	1892
Db	544	CGCGGCGCTGATTCGAGAGGCGCGCGGGGTATGCTTGCAGATACCGAGAGAGACGATCC	485
Oy	1893	TGGCAAACTGTATGATGCGCGCGGAGTCAAGCGGCTGCTGATGACGAGCGCGGCGTCA	1952
Db	484	GCAAGGCTTGGCGCGCGCGCTGAGATGTCAAGGCGCTGTGTGGCGCGCGCGCGCGCGCGC	425
Oy	1953	AGGCGAGCGGACAGCGCTCCGATGATGCCATTGACACCTTCCTGATCTTCGCGTATGT	2012
Db	424	CCCCGTGCGC-----CTGGCGCGGAGGCGGAGTCTTATGT	389

QY	2013	CATCTAACCTCCGGATCCACAGGGTTGCCAAGGGGGTGATGATGCATATCCGGGGTGC	2072
Db	388	GATTACACCTCGGGCTCCACCGGGGTGCCAAGGGGGTGCAGAGTACGCCACCGGGCGC	329
QY	2073	CGTCAACACCATCTGTGACATCAACGAGCCCTTCGAATAGGGCCCGGAGACAGGGTCT	2132
Db	328	GATCAATACCATGAGACGCGCTGCTCGACCTGCTGCGGGGTAAAGCATCGATGCTTGCT	269
QY	2133	GGGGCTTCTCTCGCTGAGCTTCGATCTCTCGGTCTATGATGTGTTGGGGATCCTGGCGGC	2192
Db	268	GGCGGTCTCCGCGCTGGACTTCATCTGTGCTTCGACTGTTCGGCGCGCTTCGGCGC	209
QY	2193	GGGGGGTACAGATGTGTGTGCGGAGCGGTCCAAAGCTGCGGATCCGGCGCATTTGGGGCAGA	2252
Db	208	CGGTGCCACGCGCTGCTCTCGCGGCGCCAGAAACAGCGCGCGATGCCGCTCTGGGGCGGA	149
QY	2253	GTTGATCGAACGAGAGAGGTGACGGTGTGGAATCTCGTGC CGGCGCTGATGCGG	2307
Db	148	GGCTATCCAGCGGATCGGGTGAAGCTGTGGAATCTGCGGCGCGGCTTGCTGGGG	94
RESULT 4			
B2562842/c			
LOCUS	B2562842	765 bp	DNA linear GSS 17-DEC-2002
DEFINITION	pac82-164_3957.y2 pac82-164 Pseudomonas aeruginosa genomic clone		
ACCESSION	B2562842		
VERSION	B2562842.1	GI:27185526	
KEYWORDS	GSS.		
SOURCE	Pseudomonas aeruginosa		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;		
	Pseudomonadaceae; Pseudomonas.		
	1 (bases 1 to 765)		
	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,		
AUTHORS	Burns,J.L., Kaul,R. and Olsen,M.V.		
	Whole-Genome-Sequence variation among multiple isolates of		
	Pseudomonas aeruginosa library		
	J. Bacteriol., (2002) In press		
TITLE	Contact: Chris K. Raymond		
	Genome Center		
	University of Washington		
	Box 352145, Seattle, WA 98105-2145, USA		
JOURNAL	Tel: 2062216954		
	Fax: 2066857244		
	Email: craymond@u.washington.edu		
	Class: Shotgun.		
COMMENT	Location/Qualifiers		
	1..765		
	/organism="Pseudomonas aeruginosa"		
	/mol_type="genomic DNA"		
FEATURES	/strain="2-164"		
	/db_xref="taxon:287"		
	/clone="pac82-164_3957"		
	/note="Clinical isolate 2-164 whole genomic shotgun library."		
BASE COUNT	157 a	248 g	125 t
	234 c		1 others
ORIGIN			
Query Match			
Best Local Similarity 3.1%; Score 133; DB 29; Length 765;			
Matches 323; Conservative 0; Mismatches 225; Indels 3; Gaps 1;			
QY	1691	GGCTGCGCGAGACAGAGGGGACGCCCGGAACATTGTGCGCGGTGTGATGAGAGAAAGCT	1750
Db	673	GTCGCGCCGAGAGAGGTGTGCGGATCAATGATCTGTCGCGGTATGCGCTGAGACACGCGC	614
QY	1751	GGAGACAGGTTGTGCGCGTCTGCGCGGTCTCGAGTACAGCGCGGCTTACGTCCGATCG	1810
Db	613	TACCCATGTGTGTGCGCTCTGCTGACGTGTCTCAAGACCGGTGCGCCTATGTCCGCTGG	554

```

QY 1811 ATGCCGACTACCGGGAGAGGATTCACCTACTCTCTGATCATGTGTAGTAAGTACG 1870
DB 553 ATCCGAGATATCCCGGAGACCGCTTGACATGATGATCAACGACGACGACGCGCTGT 494
QY 1871 TCCCTGACGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1930
DB 493 TCCCTGACGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 434
QY 1931 TCCCTGACGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1987
DB 433 TCCCTGACGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 374
QY 1988 AGACACCTTCGATCTCGGATATGATGATGATGATGATGATGATGATGATGATGAT 2047
DB 373 AGACACCTTCGATCTCGGATATGATGATGATGATGATGATGATGATGATGATGAT 314
QY 2048 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2107
DB 313 GGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 254
QY 2108 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2167
DB 253 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 194
QY 2168 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2227
DB 193 TGGAGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 134
QY 2228 TGGAGCTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2287
DB 133 CCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 74
QY 2288 CGGTGCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2308
DB 73 AGCGGCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 53

RESULT 5
AQ989512/c 646 bp DNA linear GSS 14-AUG-2000
LOCUS Rf000062 Photorhabdus luminescens strain W14 M13 library
DEFINITION Photorhabdus luminescens genomic clone pLG00062, genomic survey
sequence.
ACCESSION AQ989512
VERSION AQ989512.1 GI:9648106
KEYWORDS GSS.
SOURCE Photorhabdus luminescens
ORGANISM Photorhabdus luminescens
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Photorhabdus.
REFERENCE 1 (baaes 1 to 646)
AUTHORS f'french-Constant, R.H., Waterfield, N., Burland, V., Perma, N.T.,
Daborn, P.J., Bowen, D. and Blattner, F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium
Photorhabdus luminescens W14: potential implications for virulence
Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
JOURNAL 20378632
MEDLINE
PUBMED 10919786
COMMENT Contact: f'french-Constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bsarfc@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For
annotation of identified clones (BLASTX, BLASTN and mapping to E.
coli K12 genome) please see f'french-Constant et al. 2000, Nucleic
Acids Res.
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers
1. 646
source

```

```

/organism="Photorhabdus luminescens"
/mol_type="genomic DNA"
/db_xref="taxon:29488"
/clone="PLG00062"
/dev_stage="primary phase variant"
/clone_lib="Photorhabdus luminescens strain W14 M13
library"
/note="Genomic DNA from strain W14 was size selected (1-2
kb) and then cloned into M13 Janus."
BASE COUNT 177 a 163 c 151 g 150 t 5 others
ORIGIN
Query Match 3.0%; Score 128.8; DB 28; Length 646;
Best Local Similarity 54.7%; Pred. No. 11e-16;
Matches 256; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

1973 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2032
DB 469 CAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 410
QY 2033 CAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2092
DB 409 CCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 350
QY 2093 TCAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2152
DB 349 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 290
QY 2153 TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2212
DB 289 TCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 230
QY 2213 CGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2272
DB 229 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 170
QY 2273 TCAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2332
DB 169 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 110
QY 2333 GCCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2392
DB 109 TGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 50
QY 2393 TGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2440
DB 49 TATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2

RESULT 6
BZ548527 1117 bp DNA linear GSS 17-DEC-2002
LOCUS pacal-60_1192.s1 pacal-60 Pseudomonas aeruginosa genomic clone
DEFINITION pacal-60_1192, genomic survey sequence.
ACCESSION BZ548527
VERSION BZ548527.1 GI:27152108
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (baaes 1 to 1117)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence Variation among Multiple Isolates of
Pseudomonas aeruginosa Library
J. Bacteriol. (2002) In press
JOURNAL
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954

```

Fax: 2066857244
Email: cgraymond@u.washington.edu
Class: shotgun.
source

Location/Qualifiers
1. 1117
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pac81-60.1192"
/clone.lib="pac81-60"
/note="Clinical isolate 1-60 whole genomic shotgun library."

BASE COUNT 169 a 403 c 330 g 212 t 3 others

ORIGIN

Query Match 3.0%; Score 125; DB 29; Length 1117;
Best Local Similarity 52.5%; Pred. No. 7.9e-16;
Matches 323; Conservative 0; Mismatches 285; Indels 7; Gaps 2;

```

OY 845 AGATCCGCTCCGCGACACGAGCAATGCGTCCGCTCGGACTCTGGGGTCGATTGAAGC 904
DB 53 AGCCGCGACACCCGGCGCTTGGGATTCGACCTTCCGCGCGAGAGCGGGCTACTCGAGC 112
OY 905 GCGGTGTGCGGGAGCGCGGACTGACCCCGACGGCGCTCATCTGCGCTGATTTTCCGAG 964
DB 113 GTCTTCCGCGCGACAGTGGCGTGAACCTTGTCCAGCGTGTTCGGCTTCGCGCTTCGCGCTGG 172
OY 965 TGATGGGCGCTGAGAGCGCGAGCCCC -CGGTTTACGCTCAATATACGCTCTTCAACGG 1023
DB 173 TCTCGGCGCGCTGAGAGCGGAGGCGGCGGCAATTTCTCAGCGTCCCGTGTTCATCGG 232
OY 1024 CTCCCGCTGCATCCGCGCGTGAACGATATACCGGGGACTTACGCTCATGTGCTCTCG 1083
DB 223 CATGCCGACGACCCCGCTATCGGCGAGGTGATCGCCGACTTACCAACCTGTGCTGCTG 292
OY 1084 GACATCGACACCACTCGCGACAAAGACTTCGAAACAGCGCGCTTACGCTTATTCAGAGCAG 1143
DB 293 GAGTCCGAGATCGAGCGCGGGGTGCTTTCGCGAGGGCGGTGAAGGCTTCAGGCGAAC 352
OY 1144 CTGTGGGAAAGCATGATATCTGCGCATGTAAGCGGTATCGAGGTCCAGCGAGGCGCGC 1203
DB 353 CTCACGAGAGCATGACACACCGCCGATTCGCCGCTCGAGAGTCTCTCGCGACAGC---- 408
OY 1204 CGGGTCTGGGAGATCCACAGGAGCGGATGTTCCCGGTGTGCTGACGAGCGCGCTTAAC 1263
DB 409 --GCCCGGACAGGCGCAGCGCTCGGCGCGGTGTGCTGCGACGAACTTGGGCGAG 466
OY 1264 CAGCAAGTCTGTGTGATCACTCTGTGACAGAGGCTCGGAACTCCGCTGTACACGACG 1323
DB 467 GAGGGCTTGTCTCCGGCGGCTTTCGGCGACGCTTTCGGGATCTCAGACGAACTGTCTCG 526
OY 1324 CAGACTCTCAGCTGCTGTGATATATAGCTCTACAGACGAGATGGGGACCTGCTCTC 1383
DB 527 CAGACCCCGCAAGTGTGCTCGACACGACTCTACCGGGGTGGGAGACGATCTCTGCTG 586
OY 1384 GCGTGGGACATGCTGAGAGAGTGTTCGCCCGGACCTTTCGAGAGCAATGCTCGAAGG 1443
DB 587 GCTTGGGACAGGCTGTGCTGCTTCTCCCGAGAGCTTCCGGAACCATGTTTCAAGCC 646
OY 1444 TACGCTGTTTTTCTC 1458
DB 647 TACGTGGGGCTGCTC 661

```

RESULT 7
LOCUS B2570019 921 bp DNA linear GSS 17-DEC-2002
DEFINITION msh2_1128_Y2 msh Pseudomonas aeruginosa genomic clone msh2_1128,
ACCESSION B2570019
VERSION B2570019.1 GI:27205080
KEYWORDS GSS.

SOURCE
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
1 (bases 1 to 921)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE
Whole-Genome-Sequence Variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244

COMMENT

Email: cgraymond@u.washington.edu
Class: shotgun.
source

Location/Qualifiers
1. 921
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_1128"
/clone.lib="msh"
/note="Environmental isolate. Whole genomic shotgun library."

BASE COUNT 175 a 264 c 268 g 201 t 13 others

ORIGIN

Query Match 2.9%; Score 121.6; DB 29; Length 921;
Best Local Similarity 52.3%; Pred. No. 3.9e-15;
Matches 293; Conservative 0; Mismatches 264; Indels 3; Gaps 1;

```

OY 1574 TGCACGCGCTGTGCGCGCGCGGTGCGACAGCTGCGCATGCACTGCGCTGTGCG 1633
DB 145 TGACCGGCTGTTCAGAGAGAGGTGCGAGCCGACCGCAGCGCGCGCTGCGCTTCG 204
OY 1634 CGCGCAACACCTCACACGATCAAGCAAGCTTTCGCGCGCGCGGCTGCGGACTTGGCGCGCGC 1693
DB 205 GCGAGAGAGCGCTTGACATGCTGACGCTGAACCGCGGTGCGCAACCGCTGCGCATGCGCC 264
OY 1694 TGCAGAGACAGAGGCGACGCGCGCAACATTTGTCGCGGTGTGATGAGAAAGGCTGG 1753
DB 265 TGATGAGCGCGGATTCGCGCGCGGACCGCTGTGTGGGTGGCCATGGAAGTTTCATCG 324
OY 1754 AGCAGTTGTGCGGCTTCTCGCGGTGCTGAGTCAAGCGCGGCTTACGTCGATCGATG 1813
DB 325 AGATGTGTGTGCGCTGATGCGATCTCAAGGCGCGGCGGCTTACGTCGCGGTGAC 384
OY 1814 CCGACTTACCGCGGAGGCTATCCACTTCTCTGATCATGTGTAAGCTTAAGCTCGTGC 1873
DB 385 CGGAGTATTCGAGAGAGGCGCAGGCTTACATGTGAGAGACGCGGCTGCACTGCTGC 444
OY 1874 TGACGACGACATGCTGATGCAAACTGTATGCGCGCGCGGATTCAGCGGCTGCTCG 1933
DB 445 TCAAGCAATCGCACTTAAGCTGCGCGGTGCGGCAAGGTGTGCAAGGATTCACCTGAGC 504
OY 1934 TGAAGGAGCGCGCTCGAAGGCGAGCGGCGACAGCGCTCGGATGATGCGCATTCAGAC 1993
DB 505 AGCGGATGCTGCGGTGAGAAACATGCGGAAACATTCAGAGGAT---CGAGCTGAAG 561
OY 1994 CTTCGGAATCGCGGATATCTATCTTACACTCTGCGGATTCACAGGAGTTGCGCAAGGGGTGA 2053
DB 562 GCGAAGATCTTGCTATGTCACTTACACTTCGCTTCCAGCGGCAAGGCTCAAGGCGCGCG 621
OY 2054 TGATGATCATCGGGGTGCGCTCAACACATCTTGAACATCAACGAGCGCTTCAAGATAG 2113
DB 622 GCAACCGCATTTGCGCGCTGAGCAACCGGTTGTGTGATGACAGAGGCTTATGGGCTGG 681
OY 2114 GCGCCGAGACAGGCTGCTG 2133

```

Db 682 GCGGTTGCGACACGCTGCTG 701

RESULT 8
BZ567332 1038 bp DNA linear GSS 17-DEC-2002
LOCUS pac62-164_6886.y2 pac62-164 Pseudomonas aeruginosa genomic clone
DEFINITION pac62-164_6886, genomic survey sequence.
ACCESSION BZ567332 GI:27198451
VERSION BZ567332.1
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1038)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..1038
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac62-164_6886"
/clone_1lb="pac62-164"
/note="Clinical isolate 2-164 whole genomic shotgun
library."

BASE COUNT 187 a 330 c 304 g 216 t 1 others

ORIGIN

Query Match 2.9%; Score 120.8; DB 29; Length 1038;
Best Local Similarity 50.0%; Pred. No. 5.9e-15; Indels 1; Gaps 1;
Matches 328; Conservative 0; Mismatches 327; Indels 1; Gaps 1;

1558 CTGAGCGACATACGCTGACGCGCTGTTGCGCGCGCGGCTCAGCAGCTGCCATGCAG 1617
Db CCGGACGAGCGCGCGCTGCGACGCTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCG 139
1618 CTGCGCGTGTGTCG 1677
Db GTGCGCATTCCTCTTCCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 199
1678 CGACCTGGGGCG 1737
Db CGCTGCG 259
1738 ATGAGAAAGCGTGGAGGAGGTTGTCGCGCGTTCGCGCGTTCGAGTCAGCGCGCG 1797
Db GTGCG 319
1798 TACGTGCGCGTGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1857
Db TACGTGCGCGTGCATGCG 379
1858 GAGGTAAAGCTGCTGCG 1917
Db GCGATGCGCGTGTGCTGCG 439
1918 ATTCACGCGGTGCTGCG 1977
Db ATGCGCTGCGCGTGTGCG 499

Qy 1978 ATGCCATTACAGACACTTCCGATCTCGCGTATGTCATCTACCTCGGGAATCCAGGG 2037
Db CAGTTCGGCGCGCATCCCGAGAACTCCCTACGATGATGATACCTTCGGTTCCCGCG 559
Qy 2038 TTGCCAAGGGGGT-GATGATGATCATCGGGGTCCGTCAACACATCTCGACATCA 2096
Db CCGTCCAGGGGGTGGCGAATAGCCGAGAGAGCCGTGTGGCGCGCATACCGAGTTTCGCT 619
Qy 2097 CGAGCGCTTCGAATAGGCGCGCGGAGACAGGCTGTGGCGCTCTCTGCTGAGCTTGA 2156
Db GAGATTATTTGCGATCG 679
Qy 2157 TCTCTCGGTATGATGATGCTTCCGATCTCGCGCGCGCGCGCGCGCGCGCGCGCG 2212
Db CCGGTTGCTGAGCAGCTTACCAACGTTCCGCTTGGGTGCTCGTGTGTCTGC 735

RESULT 9
BZ548604 667 bp DNA linear GSS 17-DEC-2002
LOCUS pac61-60_1266.s1 pac61-60 Pseudomonas aeruginosa genomic clone
DEFINITION pac61-60_1266, genomic survey sequence.
ACCESSION BZ548604
VERSION BZ548604.1 GI:27152185
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 667)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
1..667
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pac61-60_1266"
/clone_1lb="pac61-60"
/note="Clinical isolate 1-60 whole genomic shotgun
library."

BASE COUNT 100 a 235 c 234 g 98 t

ORIGIN

Query Match 2.8%; Score 119; DB 29; Length 667;
Best Local Similarity 52.1%; Pred. No. 1.3e-14;
Matches 266; Conservative 0; Mismatches 245; Indels 0; Gaps 0;

2438 TGATCAGCTGCGGGCG 2497
Db TGGTCAATTTCTACCG 192
Qy 2498 ACGTGACCTATGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2557
Db TGCAGCGCGAAGCGCTGCG 252
Qy 2558 ACGTGTGATAGAGCGCGCTCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2617
Db AGTGTCTGACAGCGCTTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 312

QY	2618	GC	GGGAGGT	CGAGCT	GGCACT	GTGGGCT	ACTG	GGCGGCA	TGAAAGAA	AGCGGCA	AGACTTCC	2617
Db	313	GGG	GATCGG	CTGCG	CTGCG	CTGCT	ATTTC	CGGAC	CCGAG	AGAC	CCGCTG	372
QY	2678	TC	GTG	CA	CCCCCG	AGAC	CCG	GGAG	CGCC	CTCT	CA	2737
Db	373	GC	GTG	AT	TC	GGG	AG	CGG	CGG	CGG	CT	432
QY	2738	CC	GAT	G	AAAA	CAT	CGA	GT	T	CAT	TGG	2797
Db	433	AG	GA	CGG	CGT	GTG	GA	AA	CA	CC	T	492
QY	2798	GC	GT	T	G	AG	CT	CG	GGG	AA	T	2857
Db	493	GC	AT	G	AG	CT	G	T	G	A	G	552
QY	2858	TG	AT	T	G	T	G	T	C	C	G	2917
Db	553	GC	GC	CA	T	CT	C	A	C	G	A	612
QY	2918	AG	GC	C	A	C	G	A	G	A	C	2948
Db	613	CG	T	T	G	C	A	C	C	G	C	643
RESULT 10	B2561650	753	bp	DNA	linear	GSS	17-DEC-2002					
LOCUS	pac62-164_3367	1	pac62-164	Pseudomonas aeruginosa	genomic clone							
DEFINITION	pac62-164_3367, genomic survey sequence.											
ACCESSION	B2561650											
VERSION	B2561650.1											
KEYWORDS	GSS.											
SOURCE	Pseudomonas aeruginosa											
ORGANISM	Pseudomonas aeruginosa											
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.											
AUTHORS	1 (Baes 1 to 753)											
TITLE	Spencer, D. H., Raymond, C. K., Smith, E. E., Sims, E. E., Hastings, M., Burns, J. L., Kaul, R. and Olsen, M. V.											
JOURNAL	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library											
COMMENT	J. Bacteriol., (2002) In press											
CONTACT	Contact: Chris K. Raymond											
UNIVERSITY	Genome Center											
ADDRESS	University of Washington											
BOX	Box 352145, Seattle, WA 98105-2145, USA											
TEL	Tel: 20622316954											
FAX	Fax: 2066857244											
EMAIL	Email: ckraymond@u.washington.edu											
CLASS	Class: shotgun.											
FEATURES	Location/Qualifiers											
SOURCE	1..753											
	/organism="Pseudomonas aeruginosa"											
	/mol_type="genomic DNA"											
	/strain="2-164"											
	/db_xref="taxon:287"											
	/clone="pac62-164_3367"											
	/clone_1ib="pac62-164"											
	/note="clinical isolate 2-164 whole genomic shotgun library."											
BASE COUNT	129 a 257 c 244 g 123 t											
ORIGIN												
Query Match	2.8%	Score 117;	DB 29;	Length 753;								
Best Local Similarity	52.9%	Pred No. 3.4e-14;										
Matches 353;	Conservative 0;	Mismatches 295;	Indels 19;	Gaps 4;								
QY	2195	GC	GT	A	C	A	T	C	G	T	G	2254
Db	732	GG	GC	A	C	G	T	G	T	G	T	673
QY	2255	TG	AT	G	A	A	G	A	A	G	T	2314

Db	612	TGATTCACCCGGAAGAGGGGTGCATATACGGTGCACCTTCGTCCGGTCCATATGCTGCAGGCGCTTCT	613
Qy	2315	TCGAGCATTTTGAAGGCTGCCCCGATTCGCTTCGTAGGTCTCTCGGGCTTTTCGCTGCTGA	237
Db	612	TGCAGGATGAAAGAGCTCGCTTCCTTCGACACAGC-----TTGAAAGCATTCGTTTGCA	562
Qy	2375	GCGGCGACCTGATCCCGGTGGGCGCTGCCTGCGAGCTCCAGGCGCATCAGGCGCCGGCGTGT	2434
Db	561	GCGGCGAGGCGCCTCCCGCGGAACCGCCAGCAGCAGTCTTTCGCCAAGTCTCGCAGGCGCG	502
Qy	2435	CGGTGATCAGCTTGAGCGGGGCGGCACACCGAAGCTGTGATCTGATCCATCGGGTACCCTGTGA	2497
Db	501	GCTCTTATATACCTTATATGCGCCGACCGAGCGGGCAT-----CGACCTCACCCCATCGGA	448
Qy	2495	GGAACGTGACCTTATCGTGGCGAGCAATCCCTACGCGCGCTCCGCTGCGCAACGAGAGT	255
Db	447	CCTGCGTGAAGAGGCGAAGGACCGCGGTGCCGATTCGCGCGGCATCGCCCAACTGCGCT	388
Qy	2555	TCCAAGTCTCGATGAGGCGGCTCGAACCGCGGCCGCTCTGGGTTTCGGGGCAACTTACA	2614
Db	387	GCTACATCTCGATGGCAACTGTGAGCGGAGTGGCGGGGTGTCTCGGCAGCTGTAC	328
Qy	2615	TTGGCGGGGTGCGGGGTGGCACTGGGCTTACTGGCGGATGAAAGAGACGGCAAGACT	2674
Db	327	TGGCGGCTGCGGGGCTTGCCCTCGTGGCTACACACAGCTTCGGGGCTGACTGCCAGCGTT	268
Qy	2675	TCTCTGTGACCC--CGAGACCGGGAGCGGCTCTACAGACCGGCGATCTGGGCGCT	2731
Db	267	TCTGTGCGAGCGCCCTTCGTGTCTGGGGGCGGATGTACCGCACGGGCATTTGGCGCGCT	208
Qy	2732	ACCTGCGCGATGGAACAATCGAGTTCAATGGGGGCGTGAAGACACACCAATCAAGCTTGGCG	2791
Db	207	ACCGGCGCGAGGGGTGATCGAGTACGCGGGGCGGATGACACCAAGGTGAAGCTGGCG	148
Qy	2792	GATACCGCGTTAGCTCGGGGAAATCGAG--AAACGCTCAAGTCGATCGGAACGTACGC	2855
Db	147	GCTTGCAGATGAGACTGGCGAGATTCGAGGCAAAAGCTGTGAGCAATCCGTGGTGGCGC	88
Qy	2851	GACGCGG 2857	
Db	87	GAGGAGG 81	

```

RESULT 11
BZ558891
LOCUS
DEFINITION      BZ558891      840 bp      DNA      linear      GSS 17-DEC-2002
                  pacs2-164_109.s1 pacs2-164 Pseudomonas aeruginosa genomic clone
                  pacs2-164_109, genomic survey sequence.
ACCESSION
BZ558891
VERSION
BZ558891.1
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 840)
  Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
  Burns,J.L., Kaul,R. and Olsen,M.V.
  Whole-Genome-Sequence variation among multiple isolates of
  Pseudomonas aeruginosa library
  J. Bacteriol., (2002) In press
  Contact: Chris K. Raymond
  Genome Center
  University of Washington
  Box 352145, Seattle, WA 98105-2145, USA
  Tel.: 2062216954
  Fax: 2066857244
  Email: ckraymond@u.washington.edu
  Class: shotgun.
FEATURES
      location/Qualifiers
      1..840
      /organism="Pseudomonas aeruginosa"

```


KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 901)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
JOURNAL Contact: Chris K. Raymond
COMMENT Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: Shotgun.

FEATURES
source
1. 901
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msb2_3778"
/clone_1b="msb"
/note="Environmental isolate. whole genomic shotgun
library."

BASE COUNT 140 a 305 c 314 g 140 t 2 others

ORIGIN

Query Match 2.6%; Score 111.8; DB 29; Length 901;
Best Local Similarity 52.4%; Pred. No. 4,3e-13;
Matches 273; Conservative 0; Mismatches 242; Indels 6; Gaps 1;

Qy 1536 GCACAAACGCGACCAACGCGCTGCTGAGCAGCATACGCTGACGCGCTGTTGCGCGCG 1595
Db 601 CGCGGCGACGCGCGCGCGAGGAGGCGCTCCAGACACCTCCATGCGCTGTTGCGCGCG 542
Qy 1596 GGTGAGCAGCTGCCCATGACAGCTGCGCGTGTGTGCGCGCGAGAGAGCTCACTACGA 1655
Db 541 CGTGGCGGCGACCCACAGGCGCGCGCTGACCTTCCCGGACGACCTGAGCTATGC 482
Qy 1656 AGAGCTTTCGCGCGCTGCGCGCGACCTTGCGCGCGCGCGCGCGAGCGCGCGCGCG 1715
Db 481 CGAAGCTGATGCGCGCTCAACCGCTGCGCGCGCTGCGCGAGTCAAGCGGTGCGCG 422
Qy 1716 GAACACATGCTGCGCGCTGCTGATGAGAAAGCTGAGAGAGGTTGTGCGCGTCTGC 1775
Db 421 GGAAGTGGGGTTCGCGCGCTGCGCGCTGAGCGCTTCCGAGATGCTGCTGCGCGTGC 362
Qy 1776 GGTGCTGAGTCAAGCGCGCGCTTACGTCGCGATGCGAGCTTACCGCGCGAGCTAT 1835
Db 361 GATCTCTAAGCGCGCGCGCTTACGTCGCGAGCGCGAAATTCGCTGAGCGCGCT 302
Qy 1836 CCACACTCTCTCAATCATGATGAGTAAAGCTGCTGCTGACGACGAGCGATGCGATG 1895
Db 301 GCAATCATGATCGAGGACAGGCGGCTGCTGCTGCTCAAGCGCGCGCGCTTGA 242
Qy 1896 CAATCTGTCATGCGCGCGCGGATCAAGCGCTG-----CTCGAGAGCGCGCGCT 1949
Db 241 GGCCTCGGCGAGTGGCGCGCGAGCGCTGCGCGCTGAGTGTCTCAGAGAGAGCGCGCG 182
Qy 1950 CGAAGGCGAGCGCGAGCTTCCGATGATGCGCATTCAGACACTTGGATCTGCGCTA 2009
Db 181 GCTGACCGCGAGAGCGCGCGCGCTGCGAGCGCTGAGCGCGCGCGAGACACAGGCTTA 122
Qy 2010 TGTGATTAACCTGCGGATCCAGAGGTTGCCCAAGGGGG 2050
Db 121 CTGATTAATCTCGGCTCACCGGACGCGAAGGGGG 81

RESULT 14
B251143/c 889 bp DNA linear GSS 17-DEC-2002
LOCUS pasci-60_3008.x1 pasci-60 Pseudomonas aeruginosa genomic clone
DEFINITION pasci-60_3008, genomic survey sequence.
ACCESSION B251143
VERSION B251143.1 GI:27154724
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 889)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
JOURNAL Contact: Chris K. Raymond
COMMENT Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: Shotgun.

FEATURES
source
1. 889
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pasci-60_3008"
/clone_1b="pasci-60"
/note="Clinical isolate 1-60 whole genomic shotgun
library."

BASE COUNT 164 a 294 c 280 g 148 t 3 others

ORIGIN

Query Match 2.6%; Score 110.4; DB 29; Length 889;
Best Local Similarity 50.8%; Pred. No. 8.5e-13;
Matches 349; Conservative 0; Mismatches 322; Indels 16; Gaps 3;

Qy 2002 CTCGCTATGTCATCTACCTCGGATCCACAGGATGCCAAGGGGATGATGATCAT 2061
Db 800 CTGCTATGTCCTTCTCACTCGCGCTCCACCTTAAGCCAGGGGCGCCCAACGC 741
Qy 2062 CATCGGGTGCCTCAACCATCTCTGACATCAACGAGCGCTTCAATAGGCGCGGA 2121
Db 740 CATGGGGGTAAACCAACGCTTGTGATGACAGAGACTTATGGGCTGGCGTGC 681
Qy 2122 GACAGGTCGCGCGCTCTCTCGCTGAGCTTGCATCTCTGCTATGATGTTGCGG 2181
Db 680 ACCACGGTGTGCAAGAGACCCCTTCACTTCAAGCTGTGCTGAGCTTCTTTCGG 621
Qy 2182 ATCTGGCGGGCGGCGGAGCG-ATCGTGTGCGGAGCGCGTCAAGCTGCGGATCCGGC 2240
Db 620 GCCCTGATGACGGGGCGCTTGTGTGTACCGCGCGCGGCAACATCCGATCCGC 561
Qy 2241 GCATTGGGCAAGTTGATTCGAACGAGAAAGTACGATGTGAATCTGCGCGCGCT 2300
Db 560 GAAGCTGTGCGCTGATTCATCGGAAGGGGTGACACGCTGACATTCGTCGCTGAT 501
Qy 2301 GATCGGATGCTGTCGAGCATTTTGAAGGTCGCGCGGATTCGCTGAGTCTTCGG 2360
Db 500 GCTGCAAGCTTCTTGAAGGACGAAGAGTCGCTCTGACCAAG-----CTGAA 450
Qy 2361 GCTTTCGCTGTCGAGCGCGACTGATCCGATGCGGCTGCGCTGCGCGAGCTCAGGCGCAT 2420
Db 449 ACGCATGTTTGCAGCGCGGAGGCTGCGCGGAGCGCCAGACAGATGTTTCGCTAA 390
Qy 2421 CAGCGCGCGCTGTGCTGATCAAGCTGCGGGGCGACCGAAGCGTGTGATCTGCTCAT 2480

Db	389	GCTGCGCGAGCGCGCTCTTAAACCTTATATGGCCCCGACCGAGGGGGGCATCGACGTAC	330
Qy	2481	CGGGTACCCCCGTAGAGAACGTTCGACTCTCGTGGCCAGCATCCCCCTACGGCCGTCCGT	2540
Db	329	CCACTGGACCTGCAATGAGAGGGGACAGAGACGCGGATGCCGAT-----CGGCGGCGCAT	276
Qy	2541	GCGCACACGACGTTCCACGTCGTGATGAGAGCGCTTCGACCGGCGCCGCGTCTGGGTTCC	2600
Db	275	CGCCAACTGGGCTGTGATCATCTCTGATGGGCAACTGGAGCGCGGTGACCGGTGGGGGTCT	216
Qy	2601	GGGGCACTTACATTTGGCGGGGTTGGGCTGGCACTGGGCTTACGGCGCGATGAAGAGA	2660
Db	215	CGGGGAGCTGTAACTGGCGGGGTCCGGGGGCTTGAGGCTTACCAACCAAGCGTCCGGGGCT	156
Qy	2661	GACGCGCAGAGCTTCCCTCGTGCACCC	2687
Db	155	GACTGCGCAGCGCTTCGTGCGAGCTCC	129

RESULT 15	B2551440/c
LOCUS	B2551440
DEFINITION	633 bp DNA linear GSS 17-DEC-2002
ACCESSION	pacsl-60_3153.xl pacsl-60 Pseudomonas aeruginosa genomic clone pacsl-60_3153, genomic survey sequence.

VERSION	B2551440.1	GI:27155021
KEYWORDS	GSS.	
SOURCE	Pseudomonas aeruginosa	
ORGANISM	Pseudomonas aeruginosa	

REFERENCE	1 (bases 1 to 633)
AUTHORS	Spencer, D. H., Raymond, C. K., Smith, E. E., Sims, E. E., Hastings, M., Burns, J. L., Kaul, R. and Olsen, M. V.
TITLE	Whole-Genome-Sequence variation among multiple isolates of <i>Pseudomonas aeruginosa</i> library
JOURNAL	<i>J. Bacteriol.</i> , (2002) In press
COMMENT	Contact: Chris K. Raymond

University of Washington
Box 352115, Seattle, WA 98105-2115, USA
Tel.: 20622161954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES	Location/Qualifiers
source	1. .633

```

/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacs1-60_3153"
/clone_1bp="pacs1-60"
/notes="Clinical isolate 1-60 Whole genomic shotgun
library."
BASE COUNT      97 a      214 g      96 t
ORIGIN

```

Query Match	2.6%;	Score 109.2;	DB 29;	Length 633;
Best Local Similarity	56.1%;	Pred. No. 1.4e-12;		
Matches 229;	Conservative	0;	Mismatches 173;	Indels 6;
				Gaps 1;

2523 CCCCTACGGCGTCCGCTGCGCAACAGACGTTCCACGTGCTCGATGAGGCGCTGAACC 2582

Db 554 CCGATCGGGCCGGCGCTGGGCAACGTGGTCTGCCGCGTGTGGACGCCGAGTTCACCT 495

2583 GCGCCGGTCTGGGTCGCGGCACTCTACATTGGCGGGTCGGCTGGCACTGGCTA 2642

Db 494 GCTGCCGCTGCGCTGCCGCGAGCTGTGCATCGGCGCCTGGGCTGGCCCGGGCTA 435

2643 CTGGCGGATGAAGAGAAGACGCCAAGAGCTTCTCG-----TGCACCCCGAGACCGG 2696

Db* 434 CCTGGGCGCCCGGCGCTGAGCGCCGAGCGCTTCGTCGCCGATCCGTTCTCGGCCGCTGG 375

[illegible]

Search completed: October 3, 2003, 17:48:03
Job time : 8366.86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 4, 2003, 03:31:13 ; Search time 973.777 Seconds
(without alignments)
11088.586 Million cell updates/sec

Title: US-10-014-717-1_COPY_11872_16104

Perfect score: 4233

Sequence: 1 atgacgacatcagctctctc.....aggcgagagagctagctaa 4233

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/FCI_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/FCI_NEW_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4233	100.0	68750	13 US-10-014-717-1	Sequence 1, Appli
2	543.4	12.8	4305	12 US-10-238-075-940	Sequence 940, App
3	543.4	12.8	48715	12 US-10-238-075-921	Sequence 921, App
4	543.4	12.8	50538	12 US-10-085-959-57	Sequence 57, Appl
5	390.4	9.2	32159	12 US-10-085-959-54	Sequence 54, Appl
6	360.6	8.5	135638	12 US-10-314-657-1	Sequence 1, Appli
7	305.6	7.2	1416	9 US-09-815-242-7601	Sequence 7601, Ap
8	295.8	7.0	1875	14 US-10-156-761-3143	Sequence 3143, Ap
9	296.8	7.0	9025608	14 US-10-156-761-3184	Sequence 3184, Ap
10	270.4	6.4	5304	14 US-10-156-761-3184	Sequence 4052, Ap
11	261.4	6.2	13029	9 US-09-815-242-4052	Sequence 3629, Ap
12	246.2	5.8	11058	14 US-10-156-761-3185	Sequence 1, Appli
13	241.6	5.7	77536	12 US-09-940-3168-1	Sequence 1, Appli
14	239	5.6	88421	10 US-09-976-059-1	Sequence 1, Appli
15	233.4	5.5	5127	14 US-10-156-761-3630	Sequence 3630, Ap
16	232.6	5.5	5307	12 US-10-238-075-1078	Sequence 1078, Ap

17	230.2	5.4	1623	14 US-10-156-761-848	Sequence 848, App
18	230.2	5.4	9025608	14 US-10-156-761-3185	Sequence 1, Appli
19	221.8	5.2	7347	9 US-09-815-242-7773	Sequence 7773, App
20	218	5.2	2934	14 US-10-156-761-3186	Sequence 3186, App
21	203.8	4.8	135638	12 US-10-314-657-1	Sequence 1, Appli
22	202.2	4.8	1706	12 US-10-238-075-1079	Sequence 1079, App
23	194.6	4.6	3048	14 US-10-156-761-3146	Sequence 3146, App
24	190.2	4.5	3789	14 US-10-156-761-3185	Sequence 3185, App
25	188	4.4	3348	12 US-10-238-075-1081	Sequence 1081, App
26	177.4	4.2	4341	14 US-10-166-087-45	Sequence 45, Appl
27	177.4	4.2	32539	14 US-10-156-087-1	Sequence 1, Appli
28	161.8	3.8	1923	14 US-10-156-761-836	Sequence 836, App
29	158.2	3.7	1178	10 US-09-924-256A-91	Sequence 91, Appl
30	157.4	3.7	1545	14 US-10-156-761-1239	Sequence 1239, App
31	155.2	3.7	3471	10 US-09-974-300-2167	Sequence 2167, App
32	155	3.7	1200	10 US-09-924-256A-85	Sequence 85, Appl
33	150.8	3.6	1204	10 US-09-924-256A-87	Sequence 87, Appl
34	148.4	3.5	1803	14 US-10-166-087-43	Sequence 43, Appl
35	137.4	3.2	1172	10 US-09-924-256A-117	Sequence 117, Appl
36	136.6	3.2	3798	10 US-09-974-300-2175	Sequence 2175, App
37	136.6	3.2	3798	10 US-09-974-300-2193	Sequence 2193, App
38	132.8	3.1	1590	14 US-10-156-761-841	Sequence 841, App
39	132	3.1	3885	9 US-09-815-242-9626	Sequence 9626, App
40	126.8	3.0	3234	10 US-09-974-300-2168	Sequence 2168, App
41	126	3.0	1722	14 US-10-156-761-834	Sequence 834, App
42	122.8	2.9	4320	14 US-10-156-761-595	Sequence 595, App
43	120.8	2.9	7158	10 US-09-974-300-2171	Sequence 2171, App
44	119.4	2.8	1806	14 US-10-156-761-827	Sequence 827, App
45	118.8	2.8	3882	9 US-09-815-242-5982	Sequence 5982, App

ALIGNMENTS

RESULT 1
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Liqon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zikler, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/10/014, 717
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US/09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-10-014-717-1
Query Match 100.0%; Score 4233; DB 13; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGATCATCTGCTTCTTAAGAGCTGAGACCAAGGATATCAAGCTGGCGGCAT 60
DB 11872 ATGAGATCATCTGCTTCTTAAGAGCTGAGACCAAGGATATCAAGCTGGCGGCAT 11931
QY 61 GGGAGCGCTCTCAAGATACAGCGCCCAAGACCCCTGAACCCGAACCTCTGCTCGA 120
DB 11932 GGGAGCGCTCTCAAGATACAGCGCCCAAGACCCCTGAACCCGAACCTCTGCTCGA 11931
QY 121 ATTCGAGACCAAAAGACGATCTTAAGATGCTCCGTCAGAGACTCCCGCAGAAATCC 180

Db 11992 ATCTCCGAGACAAAAGACGATCTGAGATGCTCCGTACAGAGACTCCCCGACGAATCC 12051
Qy 181 ATCTGTCGCCGCCCAAGCCGAGCGGCGCTCCGTTTCTCTCAACAGACATCCAGAAATCC 240
Db 12052 ATCTGTCGCCGCCCAAGCCGAGCGGCGCTCCGTTTCTCTCAACAGACATCCAGAAATCC 12111
Qy 241 TACTGCTGGGGCGGACAGAGCGTTTACGTGTCGCCGAGGAGATCCACGCCATATCCGCGAA 300
Db 12112 TACTGCTGGGGCGGACAGAGCGTTTACGTGTCGCCGAGGAGATCCACGCCATATCCGCGAA 12171
Qy 301 TACGACTGTACGGATCTCCAGCTGCGAGGCTGAGCGCGGCTTTTGGAAAGTCGTCCG 360
Db 12172 TACGACTGTACGGATCTCCAGCTGCGAGGCTGAGCGCGGCTTTTGGAAAGTCGTCCG 12231
Qy 361 CGGACAGACATGCTTGGGCGCACACGCTGCGCCGACATGATGAGTATCCAGCTTAA 420
Db 12232 CGGACAGACATGCTTGGGCGCACACGCTGCGCCGACATGATGAGTATCCAGCTTAA 12291
Qy 421 GTTCGACCGCGACATCGAGATTCATCCATCTGCGCGGGCTGACCCGAGCACAGCGGAGG 480
Db 12292 GTTCGACCGCGACATCGAGATTCATCCATCTGCGCGGGCTGACCCGAGCACAGCGGAGG 12351
Qy 481 AGGCTGTGTCTTGGGAGATGCGATGTCGACCCGATCTATGACACAGCGCCCTCCG 540
Db 12352 AGGCTGTGTCTTGGGAGATGCGATGTCGACCCGATCTATGACACAGCGCCCTCCG 12411
Qy 541 CTCTATACGTGTGCGCGCTTGGGCTGAGACAGCGGCAAAACCGTCTCGTCTCACTATC 600
Db 12412 CTCTATACGTGTGCGCGCTTGGGCTGAGACAGCGGCAAAACCGTCTCGTCTCACTATC 12471
Qy 601 GATTCATTAAGTTGACCTAGGACGCTGTCATATCTTCAAGGACGCGCTCACTATC 660
Db 12472 GATTCATTAAGTTGACCTAGGACGCTGTCATATCTTCAAGGACGCGCTCACTATC 12531
Qy 661 TACGAGATCCGAGACCT 720
Db 12532 TACGAGATCCGAGACCT 12591
Qy 721 GCGCTGAGATCTCGCAAGAGTCTGAGCGGCTGACGATCGATCGATGATTACTGGAAGCG 780
Db 12592 GCGCTGAGATCTCGCAAGAGTCTGAGCGGCTGACGATCGATCGATGATTACTGGAAGCG 12651
Qy 781 CGCATGCGGAGACTCCACCTCGCGGAGGCTTCGATGAAGCGGATCCATCTTACCTG 840
Db 12652 CGCATGCGGAGACTCCACCTCGCGGAGGCTTCGATGAAGCGGATCCATCTTACCTG 12711
Qy 841 AAGGAGATCCGCTTCCGAGCACAGAGCAATGCTGCGGCTCTCTCTCTCTCTCTCTCT 900
Db 12712 AAGGAGATCCGCTTCCGAGCACAGAGCAATGCTGCGGCTCTCTCTCTCTCTCTCTCT 12771
Qy 901 AAGCGGCTGTGCGGAGAGCGCGGCTGACCCCGACGCGGCTCATCTGCTGCTGATTTTC 960
Db 12772 AAGCGGCTGTGCGGAGAGCGCGGCTGACCCCGACGCGGCTCATCTGCTGCTGATTTTC 12831
Qy 961 GAGGTGATGGGGGCTGAGCGCGAGCGCCCGGTTTACGCTCAACATTAACGCTTTCAAC 1020
Db 12832 GAGGTGATGGGGGCTGAGCGCGAGCGCCCGGTTTACGCTCAACATTAACGCTTTCAAC 12891
Qy 1021 CGGCTTCCCGCTCATCTCGCGCTGAAAGATATCACCGGAGACTTCACTGCTGATGCTCTC 1080
Db 12892 CGGCTTCCCGCTCATCTCGCGCTGAAAGATATCACCGGAGACTTCACTGCTGATGCTCTC 12951
Qy 1081 CTGAGACATGACACCATCTGCGCAAGAGCTTGAACAGCGGCTTAAGGATTTCAAGAG 1140
Db 12952 CTGAGACATGACACCATCTGCGCAAGAGCTTGAACAGCGGCTTAAGGATTTCAAGAG 13011
Qy 1141 GAGCTGTGGAGAGCGATGATATCTGCGAGCTTAAGGAGTATGAGGATCCAGGAGAGGCG 1200
Db 13012 GAGCTGTGGAGAGCGATGATATCTGCGAGCTTAAGGAGTATGAGGATCCAGGAGAGGCG 13071
Qy 1201 GCGCGGCTCTGCGGAGATCCACAGAGCGGATTTCTCCGCTGCTGCTCAAGCGCGCTT 1260
Db 13072 GCGCGGCTCTGCGGAGATCCACAGAGCGGATTTCTCCGCTGCTGCTCAAGCGCGCTT 13131

Qy 1261 AACGAGCAATGCTGTGTGTCACTCTGTTGAGAGAGGCTCGGAATCCCGGTGTACACGAC 1320
Db 13132 AACGAGCAATGCTGTGTGTCACTCTGTTGAGAGAGGCTCGGAATCCCGGTGTACACGAC 13191
Qy 1321 AGGCAATCTCCAGCTGCTGTGATCATGACTCTACAGACAGATGGGAGCTCTCTC 1380
Db 13192 AGGCAATCTCCAGCTGCTGTGATCATGACTCTACAGACAGATGGGAGCTCTCTC 13251
Qy 1381 CTGCGGTGGAGCATGTGTGACGAGAGTGTCCCGCCGACCTTTGTGAGACATGCTCGAA 1440
Db 13252 CTGCGGTGGAGCATGTGTGACGAGAGTGTCCCGCCGACCTTTGTGAGACATGCTCGAA 13311
Qy 1441 GGTACGTGCTTTTCTCCGCGGCTCACTAGAGAACATGGGAGTGAACAGGTGCGCTGT 1500
Db 13312 GGTACGTGCTTTTCTCCGCGGCTCACTAGAGAACATGGGAGTGAACAGGTGCGCTGT 13371
Qy 1501 TCGCTTCCGCTGCCAGCTAGAAAGCGGAGCGGAGAAACGCGACCAACGCGCTGCTG 1560
Db 13372 TCGCTTCCGCTGCCAGCTAGAAAGCGGAGCGGAGAAACGCGACCAACGCGCTGCTG 13431
Qy 1561 AGCGAGATACGCTGACAGGCTGTTGCGCGCGCGGCTGAGACAGCTGCCATGCACTC 1620
Db 13432 AGCGAGATACGCTGACAGGCTGTTGCGCGCGCGGCTGAGACAGCTGCCATGCACTC 13491
Qy 1621 GCGCTGTGTGTGGCGCGGAGAACGCTCACTACGAAAGCTTTGCGCGCGCTTCCGCGGGA 1680
Db 13492 GCGCTGTGTGTGGCGCGGAGAACGCTCACTACGAAAGCTTTGCGCGCGCTTCCGCGGGA 13551
Qy 1681 CTGGCGCGGCTCGGAGACAGGAGGCGAGCCCGAACAATGCTGCGGCTGTGTATG 1740
Db 13552 CTGGCGCGGCTCGGAGACAGGAGGCGAGCCCGAACAATGCTGCGGCTGTGTATG 13611
Qy 1741 GAGAAAGCTGGAGACAGGTTGTGCGGTTCTCGCGTGTCTGAGTCAAGCGCGGCTTAC 1800
Db 13612 GAGAAAGCTGGAGACAGGTTGTGCGGTTCTCGCGTGTCTGAGTCAAGCGCGGCTTAC 13671
Qy 1801 GTTCGCGATGATCCGACCTTACCGGCGGAGGTAATCCATCTCTCTGATCATGATGTAG 1860
Db 13672 GTTCGCGATGATCCGACCTTACCGGCGGAGGTAATCCATCTCTCTGATCATGATGTAG 13731
Qy 1861 GTTAAAGCTGTGCTGACAGCAGCATGCTGTGATGCAACTGTCAATGCGCGCGGAGATC 1920
Db 13732 GTTAAAGCTGTGCTGACAGCAGCATGCTGTGATGCAACTGTCAATGCGCGCGGAGATC 13791
Qy 1921 CAGCGGCTGTCTGTAGCGAGGCGCGGCTCGAAGCGACAGCGGATCCAGCTTCCATGATG 1980
Db 13792 CAGCGGCTGTCTGTAGCGAGGCGCGGCTCGAAGCGACAGCGGATCCAGCTTCCATGATG 13851
Qy 1981 CCCATTCAAGACCTTCCGATCTCGCGTATGCTCACTCACTCGGAGTCCACAGGTTG 2040
Db 13852 CCCATTCAAGACCTTCCGATCTCGCGTATGCTCACTCACTCGGAGTCCACAGGTTG 13911
Qy 2041 CCCAAGGGGTGATGATGATCATCGGGGTGCGGCAACACCATCTCGAGATCAACGAG 2100
Db 13912 CCCAAGGGGTGATGATGATCATCGGGGTGCGGCAACACCATCTCGAGATCAACGAG 13971
Qy 2101 CGCTTGAATTAAGGCGCGGAGACAGGTTGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 2160
Db 13972 CGCTTGAATTAAGGCGCGGAGACAGGTTGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 14031
Qy 2161 TCGCTTATGATGATGTTGGGATCTCTGCGCGGCGGGAATGATGATGCTGCGGAGCGG 2220
Db 14032 TCGCTTATGATGATGTTGGGATCTCTGCGCGGCGGGAATGATGATGCTGCGGAGCGG 14091
Qy 2221 TCCAAAGCTGCGGATCCGCGGATTTGGGACAGATTGATCGAAGAGAGTGAACGCTG 2280
Db 14092 TCCAAAGCTGCGGATCCGCGGATTTGGGACAGATTGATCGAAGAGAGTGAACGCTG 14151
Qy 2281 TGAAGCTGCTGCGCGGCTGATGCTGCTGCTGAGCATTTTGAAGGTTGCCCGAT 2340
Db 14152 TGAAGCTGCTGCGCGGCTGATGCTGCTGCTGAGCATTTTGAAGGTTGCCCGAT 14211

2341 TCGCTGCTAGGCTCTCTGCGGCTTCTGCTGCTGACGCGGCACTGATCCCGTGGGCTTG 2400
Db 14212 TCGCTGCTAGGCTCTCTGCGGCTTCTGCTGCTGACGCGGCACTGATCCCGTGGGCTTG 14271
Qy 2401 CCTGGCGAGCTCCAGGCGCATGAGGCGCGGCTGCTGCTGATCAAGCTTGGGGGCGGCGAC 2460
Db 14272 CTTGGCGAGCTCCAGGCGCATGAGGCGCGGCTGCTGCTGATCAAGCTTGGGGGCGGCGAC 14331
Qy 2461 GAAGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520
Db 14332 GAAGCTGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14391
Qy 2521 ATCCCTTACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
Db 14392 ATCCCTTACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14451
Qy 2581 CCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640
Db 14452 CCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14511
Qy 2641 TACTGCGCGATGAG 2700
Db 14512 TACTGCGCGATGAG 14571
Qy 2701 CGCCTTACAAAGACCGCGGATCTGCGGCGCTACCTGCGGATGAGAGAGAGAGAGAGAGAGAG 2760
Db 14572 CGCCTTACAAAGACCGCGGATCTGCGGCGCTACCTGCGGATGAGAGAGAGAGAGAGAGAGAG 14631
Qy 2761 GGGGCTGAG 2820
Db 14632 GGGGCTGAG 14691
Qy 2821 GAAAGCTCAAGTGCATCCGAG 2880
Db 14692 GAAAGCTCAAGTGCATCCGAG 14751
Qy 2881 GCGGCGAG 2940
Db 14752 GCGGCGAG 14811
Qy 2941 GAG 3000
Db 14812 GAG 14871
Qy 3001 GACGCTTACGAG 3060
Db 14872 GACGCTTACGAG 14931
Qy 3061 GACCTGAG 3120
Db 14932 GACCTGAG 14991
Qy 3121 GACGCTTACGAG 3180
Db 14992 GACGCTTACGAG 15051
Qy 3181 GAGTTTGGTTCGATTCCTGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Db 15052 GAGTTTGGTTCGATTCCTGAGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15111
Qy 3241 AAATTCCTTATTCATCGGCGGAG 3300
Db 15112 AAATTCCTTATTCATCGGCGGAG 15171
Qy 3301 TCCGCGCGCATCGAGGCGGTGAG 3360
Db 15172 TCCGCGCGCATCGAGGCGGTGAG 15231
Qy 3361 TTGCTGAAGCTCTCGATCAAG 3420
Db 15232 TTGCTGAAGCTCTCGATCAAG 15291
Qy 3421 GTGTTTGATGAAGCGGCGTTCAACTCTGTTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480

15392 GTGTTTGATGAAGCGGCGTTCAACTCTGTTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 15351
Db 15392 GTGTTTGATGAAGCGGCGTTCAACTCTGTTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 15351
Qy 3481 CTGATGAGATGCTGCTGAG 3540
Db 15352 CTGATGAGATGCTGCTGAG 15411
Qy 3541 CTGATGAG 3600
Db 15412 CTGATGAG 15471
Qy 3601 GAACAGGTTGCGGCGGCTTCTGAG 3660
Db 15472 GAACAGGTTGCGGCGGCTTCTGAG 15531
Qy 3661 GGGGCGGAG 3720
Db 15532 GGGGCGGAG 15591
Qy 3721 AGCGCGCGCAG 3780
Db 15592 AGCGCGCGCAG 15651
Qy 3781 CGGCACTTCTTGAAG 3840
Db 15652 CGGCACTTCTTGAAG 15711
Qy 3841 GATGCGTTGCGGCGAG 3900
Db 15712 GATGCGTTGCGGCGAG 15771
Qy 3901 GATGCTGCTGCGCGCGGAG 3960
Db 15772 GATGCTGCTGCGCGCGGAG 15831
Qy 3961 CTGCTGCGGCTGCTGAG 4020
Db 15832 CTGCTGCGGCTGCTGAG 15891
Qy 4021 GTGATCTGCTGCTGAG 4080
Db 15892 GTGATCTGCTGCTGAG 15951
Qy 4081 CTGATGAG 4140
Db 15952 CTGATGAG 16011
Qy 4141 TCCGCTTTCGCGAG 4200
Db 16012 TCCGCTTTCGCGAG 16071
Qy 4201 GAGGTTTTCGCGAG 4233
Db 16072 GAGGTTTTCGCGAG 16104

RESULT 2
US-10-238-075-940
Sequence 940, Application US/10238075
Publication No. US20030148324A1
GENERAL INFORMATION:
APPLICANT: I.N.S.E.R.M.
TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
TITLE OF INVENTION: E.coli, and biological uses of these polynucleotides and of their
FILE REFERENCE: BLANDINE
CURRENT APPLICATION NUMBER: US/10/238, 075
PRIOR FILING DATE: 2002-09-10
PRIOR APPLICATION NUMBER: 0003145
NUMBER OF SEQ ID NOS: 1576
SOFTWARE: Patent version 3.1
SEQ ID NO 940
LENGTH: 4305

2257 CTGCGGAGACGACGCGGAAAAAGATCCGCGTCAATTGGCAGACATTAATGCGACACGGT 2316
2269 AAGTGAAGTGTGAGACTCGGTGCGCGGCTGATGCGGATGCTGCGAGCATTTTGG 2328
2317 CATGTAAGCTGTGATGAGATGCGCGGCTGATGAGATGCTGTGG---AATACAC 2373
2329 GGTGCGCCCATTCGCTGCTGATGATGCTGCGGCTTTGCTGCTGAGCGGCGCATGTATC 2388
2374 AGCGCGATCGGATGATGATTCGACGATGCTGCGCATGTTGAGCGGCGCATGTATC 2433
2389 CCGGTGGGCTGCTGCGGCGA---GCTCAGGCGCATAGGCCCGGCGTGTGATGAC 2445
2434 CCGCTAAGTTACCGGAGCAGATGCGCAGCGGCTCAATGAACCATGACATCAATCAAT 2493
2446 CTGCGCGGCGGCGACCGGAGCGCTGATGCTGCTGATGCTGCGGATCCCGTGAAGACGTCAC 2505
2494 CTGGGTGAGACGACCGGATGCGGCGCATGCTGCTGCTGATGCTGATGCTGATGAGTGA 2553
2506 CTATGCTGCGGCGGATGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCTG 2565
2554 TCGAGTGGAGCAGATGATTCCTTACGCTGCGGCGCTGCGGCGGCTGCGGCGGCTGCTA 2613
2566 GATGAGGCGCTGAGACCGCGCGCGGCTGCGGCTGCGGCGGCTGCTGATGCTGATGCTG 2625
2614 AATGCGCAATGAGAGAAATGCTGCTGCGGCTGAGAGAGATTTGATGCTGCGGATG 2673
2626 GGGCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2685
2674 GGGCTGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2733
2686 CCGGAGCGGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2745
2734 GAGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2793
2746 AATCATGAGTTCATGCGGCTGAGAGACCAACCAATCAATCTGCGGATGCTGCTGCTG 2805
2794 CAAGTCTGCTTTTGGGCGGCGCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2853
2806 CTGCGGAGAAATGAGAGAAAGCTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2865
2854 CTGGGAGAAATGAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2913
2866 CCGCTGCGGAGCGAGCGGCGGAAAGCTCTTCTGAGCTGATGCTGCTGCTGCTGCTG 2925
2914 GCAGTGGGTA-----TTCTCAATCATGCTGG 2940
2926 CCGAGACCGCGCTGCGGAGACGAGCGCTCAAGACCGAGCGGATGAGCGGAGAC 2985
2941 CTGCTGCTTTTTCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3000
2986 CACGCGCGGAGCGAGCGGCTTGAAGCGAGCGGAGGCTGCTGCTGCTGCTGCTGCTG 3045
3001 -----GCGAGCGGCGGCGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3054
3046 CACGAGCTCGAGGAGCTGAGCGGAGAAAGCGGCTGCTGCTGCTGCTGCTGCTGCTG 3105
3055 GCGTTCAATCTCAAGAGGACATATGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTG 3114
3106 CCGAGGCGGCGGCTGAGCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3165
3115 GCACCGGCGGAGTACGACATGCTCAATCAACCGCGGCTGCTGCTGCTGCTGCTGCTG 3174
3166 CCGATTCGCTTTGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3225
3175 AAAACACGCTGCGACATGTTGGGAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3234
3226 GCGCGACCTTCCCAATTCGCTTATCATGCGGCGGAGACGCTACCGGCTGCAAC 3285
3235 GGTCACTTTTTCGCAATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3294
3286 TACGCTATGCTCAATTCGCGCGGCTGAGGCGGCTGAGAGAGGCTTCTATTTATTCAC 3345
3295 TACGTATCTGCTGCTGAGACAAAGTCAAGAGGAGATATCGGATTTACTTACTTGCAC 3354

3346 CCGTTGAGACCGCTTGTCTGAGCT---CTCCGATCAGGAGATGAGCGGCGGAC 3402
3355 CCGGAGAGAGCTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3414
3403 GTTGGCAAACTTTCAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3462
3415 GCGGAGCTTATCACTGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3474
3463 ATGACGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3519
3475 ATGCGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3534
3520 GCGGATATATGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3579
3535 GCAGTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3594
3580 CCGGTGGGCAATTCATTTTGAACAGGTTGCGCGGCTTCTGACCTGCTGCTGCTGCTG 3639
3595 CACTTGGGATGAGGAGCTTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3654
3640 GTTATGCTGAGCGGATGCTGCGGCGGCGGCTGAGACCGGCGGAGTTCCAGGCTGCTG 3699
3655 CCGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3714
3700 CTGCTGAGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3759
3715 TTGCGGCGGATTTCTTCCACTATGAGAAACGAGAGAAACGCTGCGGCGCTGAGCATG 3774
3760 CAGCATCTTCCGCTATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3819
3775 CAGAGCTTACAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3834
3820 CCGTACGCTTCTGAGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3879
3835 CCGAGTATCTGATGCTTACGACGATTTTCTGCTTACCGCTTACCGGCTGCTGCTGCT 3894
3880 AAGCGCTGCGGAGCGGAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3933
3895 CAATAATTCAGCTGCGAGGCGGAGCAATGCTCCACCGGCTGAGCGGCTGCTGCTGCTG 3954
3934 GCGCGACGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3993
3955 CAGGTGAGAGTGTGCTTACCAACAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4014
3994 GAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4053
4015 AGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4074
4054 CCGATGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4113
4075 CGTATTCAGAGGAGCACTTGAAGCGATTAATCGGAGAGATTCCTGCTGCTGCTGCTG 4134
4114 CAGTACCGCAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
4135 GGTCTGCCAACCATCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4161

RESULT 3
US-10-238-075-921/c
; Sequence 921, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I. N. S. E. R. M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolat
; FILE REFERENCE: Blandine
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 0003145
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1

SEQ ID NO 921
LENGTH: 48715
TYPE: DNA
ORGANISM: Escherichia coli
US-10-238-075-921

Query Match 12.8%; Score 543.4; DB 12; Length 48715;
Best Local Similarity 48.5%; Pred. No. 1.1e-131;
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;

181 ATGTGTCCTCCGCGCCGAGCGGCGGCGCTGCTGCTCTCTCAAGACATCCAGATCC 240
18373 ATAGTGACAGACCTGAGCATGTATCAACGGTTTCCATTGACGATTTTCAAGCGCCG 18314
241 TACTGGCTGGCGCGGACAGAGCGTTTACGTTCCAGCGGATCCAGCCTTATGCGGA 300
18313 TACTGGCTTGGAGC--ACAGACGGGGTGTACTCTCATCGCAGCCACATCTACCATGAA 18257
301 TACGACTGAGGATCTCGACGTCGCGGAGCTGAGCGCGGCTTTGGAAGTCGCGG 360
18256 TTTGACGTAGAAACATTATATGTATGCGCTTTTACCATGCGGTGAATCGCTGATCGCT 18197
361 CGGACAGCATGCTTCCGCGCCACACGCTGCCGACATGATGACGAGTATCGACCTTAA 420
18196 CGCCATGAAATGTCTAGTGGCGGGGATCTCCCGACGATACACAGATTTCTGGCGCA 18137
421 GTGACGC--CGACATCGAATCATGATCTGCGCGGGCTGACCGGAGCACAGGGAA 477
18136 GTGCGCGCTATCATGTTAGAGAGCGCGATGAGGCTTTGCTCCCTAACGACGAAAC 18077
478 GCGAGCGCTGTCGCTGTTGGAGATGAGATGCGACCGCATCTATGACACGAGCGCCT 537
18076 GATGCTTATGAGCATCCGCGATCGGCTGTGCGATCATGTGATCCGGAATCTTGG 18017
538 CGGCTCTATCAGCTGCTGCGCGTGGCTGAGACGAGCGGCAACCCGCTCTGCTCACT 597
18016 CGGCTGTTGATTGATTGATTATTCGCTTGCACGCGGCAACATGGCGGCTTGATTTCACT 17957
598 ATGATCTCATTAAGTTAAGTAACTAGGACGCTGTCCATCATCTTCAAGGATGCTCAGC 657
17956 CTCGATCTGCTCATATGCGCATCTGATGATGCGGACGCTACAGCAGAGATTGATGAG 17897
658 TTTCTGCAAGATCCGCGAGCTCTCTCCGTCCTGAGCTCTGATCCGACGCTATATGTA 717
17896 CTGTACCTGAGGCCCATATGTGTACTGCGCTGTCTACCGTTCTCTTTTGTGACTAGTG 17837
718 CTCGCGCTGAGATCTCGCAAGAGTCTGAGCGCATCAACGATCGATGATTTACTGAG 777
17836 CAGCGGCTGTTGGTGAAGCAGGCGGATGAAGCCTATGCAAGCGATCAGGCTATTGGCA 17777
778 CCGCGCATGCGGAGCTCCCACTCTCGCGAGCGTTCCGATGAAGCGCATCTTACC 837
17776 CCGGCGCTCCGACGCTGTATGAGCCCAACAGCCTGCGCTACAGGCGCATTTGGCGCA 17717
838 CTGAAAGGATCGCTTCCGCGACACAGAGCAATGCTGCGGATCGGACTCTGGGGTGA 897
17716 CTGTCTGCGATCAGGTTCTGTAGTGTGCGGCTCATCGCTGTCAACCCACAACTGGGAGTG 17657
898 TTGAAGCGCGCTGTGCGGAGCGCGGCTGACCCCGACGAGCGCTCATCTGTGCTGATTT 957
17656 CTGAGCGGCTGCGCCACAGCAACAGTATCACCAAGACGGAATGTTGTTGACAGTCTTT 17597
958 TCCGAGGTATGCGGCGCTGAGACGCGAGCGCCCGGTTTACGTTCAACATTAACGCTTTC 1017
17596 ACGCAAGTCTGCGAGTTGAGGCTTACCGCGACCTTTACGTTCAATCTGACGTTGTC 17537
1018 AACCGGCTCCCGTCCATCGCGCGTGAACGATATACCGGGGACTTCACTGTCGATGTC 1077
17536 AACCGCGCGAGGTTTACCCCAACGAGAGGAGTTATGTTGATTTTACCGCTGTACG 17477
1078 CTCCTGAGATCGACACGCTCGCGCAAGAGCTTGAACAGCGCGCTAAGCGTATTCA 1137
17476 TTGCTGAATGTTTGTTAACGACAGCGACGCTTTATGCGCCACAAAGCTCAGGCTATTG 17417

1138 GAGCAGCTGTGGGAAGCATGATGATACCTGCGACGTAAGCGGAT-----CGAGGTC 1188
17416 GTGCACTGTGGGAAGATCTGAAACATGCTGTTTCATGTGGATCCGCCACGACGAGCG 17357
1189 CAGCAGAGCGCGCGCGGCTCTGGGAGATCCAAACGAGCGGATTTTCCCGTGTGTC 1248
17356 CTGATCCATTAAGCGCTGTTTCATATGCGCGGATCCCGGTGTATTCATAGTATGTTGAT 17297
1249 ACGAGCGGCTTAAACGAAAGTGTGTGTCTCATCTGTTGACAGAGCTGGAACCTCG 1308
17296 ATCGACGGGAGAACACTGCGCAAGACCTCGGAGACAAACCGTTTACTGTGTGTCG 17237
1309 GTGTACACAGACGACATCTCTAGCTGCTGATCATACGCTGATCGAGACGAT 1368
17236 GACGCAATATTATCCCAAAACCGAGGTGTGCTGATACACAGATATGAGTTGCT 17177
1369 GGGAGCTGTCTGCTGCGTGGGACATCTGTCAGAGAGTTTCCCGCCGACCTTGTGAC 1428
17176 GGGAGTTGCACTTCACTGGGACCGGCTGAGCACTGTTGATATCACGCTGTGGAT 17117
1429 GACATGCTGAAAGCTGACGTGTTTCTCCGCGGCTCACTGAGAAACATGGGTGA 1488
17116 CAGATGTTGGCTTATTGTATGCTGACGCTGACAGCGGCTGTGTCATGCC-----G 17066
1489 CAGGTGCGCTGTGCTTCCGCTGCCCTGCCAGCTAAGAGCGCGGCGCAAGCGCAACC 1548
17065 CAAGTTGTGTGGGGGTAAATAGTCTGTGGCGCTGCCACCGTTATGTGACCGGTACG 17006
1549 AACGCGCTGTGAGGAGCATATCGCTGACGCGCTGTTGCGGCGCGGCTGAGCAGTGG 1608
17005 CAGGCTCTGACACCTACCGGCTGTGTTGACATGATATTAAGTGTGACGAGCACTGACG 16946
1609 CCATGACGCTGCGCTGTGTGCGCGCAAGACGCTACAGTACGAAGCTTTGCGGC 1668
16945 CCACAGAACTGCGCTGATAGTCTATCCGTGAATTGACCTATGCGCACTGTGACG 16886
1669 CGTTGCGCGGACATTTGGGCGCGCGGTGCGGACAGAGGGGGACCGCCGAACATTTGTC 1728
16885 GCGGCGAATCATGTGCGCGCGCGCTGTTAGCGCTGAGCGGTGACATGCGACCGCGTG 16826
1729 GCGGTGTGATGAGAAAGCTGGAGAGGATGTGCGGCTTCTGCGGTCTGAGTCA 1788
16825 GCGGTGTGATGAGAAAGCTGGAGAGGATGTGCGGCGGTGACATGCGGATTTTACGATG 16766
1789 GCGCGGCTTACGTGCGCATGATGCGACCTACGCGGAGCGTATCTCACTCTCTC 1848
16765 GGTGCGGTCTATCTGCGCATGATGCGGTGCTACCGCACAGCGCTGCGAGCTTTGCTG 16706
1849 GATCATGATGAGGTAAAGCTGTGCTGACGCGACGCGATGAGGAGCAACTGTGCATGG 1908
16705 ACGGTGGGAGGTGCGGTACAGTAAAGCAGCGGATCTACACCAATTGAGGCGGTGCG 16646
1909 CCGCGCGGAGATCCAGCGGCTGCTGTGACGAGCGCGGCTGAGAGCGACGCGACGAG 1968
16645 CTGCGCG-----TGTGATCATGACGAGGAATGCTGACACGCGCTGCTGCG 16598
1969 CCTCCGATGATCCCATTCAGACACTTGGATCTGCGGTATGTATCTACACTCTGGGA 2028
16597 CGTTGCTGAAAGTGTGCGGTGAGATGACGAGATCTGCGGCTTATATATTCTTCCTCGCG 16538
2029 TCCAGAGGTTGCCAAGGGGTGATGATGATCATCTCGGGGTGCGCTCAACATATCTG 2088
16537 TCCACCGGATCCCAAGAGATGATGATGACACCGCTGCGGCTATGACACGCTGGA 16478
2089 GACATCAAGAGCGCTTGAATAGGCGCGGAGACAGGCTGCTGCTCTCTGCTG 2148
16477 GACATCAAGAGCGCTTGTGCTGCTCAATGCGAGATGAGGATGTTGGGCTGTATCATTTG 16418
2149 AGCTTGATCTCTGCTGTATGATGTGTTGCGGATCTGAGCGGCGGCTGATCATCTG 2208
16417 AGCTTTGACCTGTGCTTATGATGCTCTTTGCGCTTTTATGTTGTTGAGGCTGACGCTGTGA 16358

QY 2209 GTGCCGACGCGTCCAGCTGCGCATCGCGCATTTGGCAGAGTTGATCGAACGAG 2268
 Db 16357 CTGCGGAAGACGAGACGGAAAAAGATCCGCGTCACTTGGCAGACAGTTATGGCACACG 16298
 QY 2269 AAGGTGAGGTGTGAACTCCGTGCGCGCTGATGCCGATCTCTGTGAGCATTTTGA 2328
 Db 16297 CATGTAACTGTGAATGCAAGTGCCTGACATGACAGATGCTGTGG---AATACAC 16241
 QY 2329 GATGCCCGCGATTCGCTGAGTCTCTGCGGCTTTGCGTCTGAGCGGCGCATGTGATC 2388
 Db 16240 AGCGCGCATCGAGTATGATTAATCCGACGTGGCTGTGGCACTGTTAAGCGGCGCATGGATC 16181
 QY 2389 CCGGTGGGCTGCTGCGGCA---GCTCCAGGCATCAGACCGCGGCTGTGCGTATCAGC 2445
 Db 16180 CCGCTAATGTTAATCCGAGCAGATGCGCGCATCAATGAAACATGAGACATCACTAGT 16121
 QY 2446 CTGGCGGGGCGCACCGAAGCGTCCATCTGTGTCATTCGGGTACCCGTGAGAACTGCAC 2505
 Db 16120 CTGGGTGAGGCGACCGAGTGGCCCATCTGTGCTACTACCCGATAGTGAAGGTGAA 16061
 QY 2506 CTATCGTGGGCGAGATCCCTTACGCGCGCTCGCGCAACAGACGTTCCAGCTGCTC 2565
 Db 16060 TCGAGTGGACCAAGATTCCTTACGCTGCGGCGCTGCGCAACGACGATATACGCTTA 16001
 QY 2566 GATGAGGCGCTGCAACGCGCCCGGTCTGTGCTTCCGGGCGCACTCTACATTGGCGGGTC 2625
 Db 16000 AATGGCAATGAGAGATGTCGCGTCCGGGTGAGAGAGATTTGATTTGGCGGGATG 15941
 QY 2626 GGGCTGTGCACTGGGTACTGCGCGCATGAGAGAAAGACGCGCAAGCTTCTCTGTGAC 2685
 Db 15940 GGGCTGGGCAAGGCTTACTGACAGACGAGAGAAACGCGGCGCAAGCTTGTCTGGGGC 15881
 QY 2686 CCCGAGACGGGAGGCGCCTCTACAGAGACCGGCGACTTGGGCGCTTACCTGCGGATGA 2745
 Db 15880 GAAAGCATGTGTGACGCAATTTACCGCATGCGGATTCGCGGCGCTTACTTGTGACGAG 15821
 QY 2746 AACATCGATTATGCGGCGTGAAGCAACCAATCAAGCTTCCGCGATCCGCGTTGAG 2805
 Db 15820 CAAGTCGCTTTTGGGCGGCAAGATCCCAATGAGAGTGAATGGTTACCGTATGAA 15761
 QY 2806 CTGCGGGAATTCGAGAAACGCTCAAGTCGATCCGAACGTAACGCGGCTGATTTTG 2865
 Db 15760 CTGGGGAATTCGAGCGCTGATTCGCGACATCCGATGTGAGACATCAAGTGTGTG 15701
 QY 2866 CCCGTCGGGAAAGACGCGGCGCAACAGCTCTTACGCTATGTGTCCCGAGGGGACA 2925
 Db 15700 GCAGTGGGTA-----TTCTCAACATGTCGG 15674
 QY 2926 CGAGACGCGCTGCGGACAGACGAGCCTCAAGACCGAGCGATCGACCGAGACA 2985
 Db 15673 CTGGTCGCTTTTGGCAATCTGACATGCGCACAGGGCGGCGATTCGCAAGCTAAGAA 15614
 QY 2986 CACGCGCGGAGCGGCTTGAAGCAAGCGGAGGCTGCACTTCAAGTCTGCTGCA 3045
 Db 15613 -----GGGAGGCGCGGCACTGCGGAGGTATTAATGTGAATCCGCAACAGCTCTA 15560
 QY 3046 CACGAGCTCCGAGGAGCTTGAAGGAAAGCCGTCGATCTGAACGGGAGGAGATCGG 3105
 Db 15559 GCGTTCAATCTCAGAGGACATATTCGCGCGCTGAGATGGTGGGCAATTCGACTGACG 15500
 QY 3106 CGGAGGCGGGGCTGAGAGTCTACGCGGCTGCGCTTACGCTCGAAGCTTCTTGAAGCC 3165
 Db 15499 GACCGGGGAGTAGACACGTTACATCAAAACCGCGACGTAATTCGATTCAGCGCGAA 15440
 QY 3166 CCGATTCCGTTTGTGATTTGTGCTTCTGACCTGCTTGAAGCAGTGAAGCCGAC 3225
 Db 15439 AAAACCAAGCTGCGCACAGTGGGGAATTCGCTGTGGGCTTGGGCGACATGCTTACCC 15380
 QY 3226 GAGCGGACCTTCCCAATTCGTTATCCATCGCGGCGAGACGTAACCGGTGAAC 3285
 Db 15379 GGTCTACCTTTTGGCAAGATGCTATGCTGCGCGGGGCTATACCGGCTGAAC 15320
 QY 3286 TACGCTATGTCAAATCCGCGCGCATCGAGGCGTGAAGAGGCTTCTATTAATACAC 3345

Db 15319 TACGTATACCTGATCCAGACAGATCGAAGAGGAGATCCGATATTACTACTTCAC 15260
 QY 3346 CCGTTGACGACCGTTTGTGAACT---CTCCGATACCGGGATCGACGGGAGCGGAC 3402
 Db 15259 CCGGACAGAGCTTCTTATGCGGTAGCACCAAGTGAAGTGAAGAGTGTTCAT 15200
 QY 3403 GTTGGCGAAACTTCGAGCTGTGATGAAGACGGCGTTCACACTCTGTGGGCGAG 3462
 Db 15199 GCCGACCTATACGTCTATTTCCGATCGGCGGCACTTACGCTGTTATATGTGCTGAT 15140
 QY 3463 ATGACCGCATGAGTGTCTGATG---ATGCTGTGCGAAGATTTTGCCTGTGAG 3519
 Db 15139 ATGCGGATCTGCGCATTTATGCGCAAGAGCAGCTTGACCTTCCGCTGATGAA 15080
 QY 3520 GCCGATATATGAGGAGCTCTGATGAGAGGCGCTTCTCGCAACATGGGCTGCT 3579
 Db 15079 GCAAGTACTCTGCTCATTTACTGAAAGATGCGCGCTACGATGGGCTGTG 15020
 QY 3580 CCGGTGGGCAATTCATTTTGAACAGTTCGCGGCTTCGACCTGCAATTCGAGC 3639
 Db 15019 CAATTTGGAGATGAGACTTTTCCGCTGTGCAATGCAATTTCAATGTCCGACATCAT 14960
 QY 3640 GTTACGTGACGCGCATGCTGGCGGCGGTGACCCGCGCACTTCAGCTCTGACG 3699
 Db 14959 CGTATGTCCATTCGACCGTGGGCGCGCATAGGCGCAAGGCGGCAAGTGTGACGA 14900
 QY 3700 CTGCGTCAAGATTCCTACAGAGCGCGCAACGCGCGGCGCCCTCCGCGCGGAG 3759
 Db 14899 TTGCTGCGCGATTTCTCACTTATGAGAAACCGAAGAAACGCTGCGCTGGACATG 14840
 QY 3760 CAGCACTTCGCGATATGCTTCGCGCATTTTGAAGCACTTACCGAGTATCATGCTG 3819
 Db 14839 CAGAGCTCAAAAGATGCACTCTGCGGCTGCTGCAAGCACTGCTGATATATGCTG 14780
 QY 3820 CTTACAGTCTGTGTGAGCTCGATGCTGTGCGCTGACGTCACACGCAAGCTGATGCT 3879
 Db 14779 CCGAGTATCTGATGTTAGCAGACGATTTTCGTTAACCGGTAACGCGCACTGATGCTG 14720
 QY 3880 AAGCGCTGCGGCGGAGAGATACCTCTGCTGCGCGGCACT-----TCGGGCGACG 3933
 Db 14719 CAATAATTCAGCTGACAGGCGGAACTTCCACACGCTGACGCGGTGGTCAATC 14660
 QY 3934 GCGCGACGAGCGCTTGAAGAGATCTCTGCGCGTGTGACGAGAGTCTGCGGCTG 3993
 Db 14659 CAGGTGACAGTGGCTTACACACGCGCTGTGCGCTTGGCAGAGATCTGCGCTG 14600
 QY 4054 CGCATGAGAGCTTGTGCAAGAGAGCTGATAGGAGATCCGATCAACGAGTGTTC 4113
 Db 14539 GGTATTCAGACAGGACATTCGAGGCGATTAATCGGAGAGATTCCTATGTTGATCTGTC 14480
 QY 4114 CAGTACCGAATCTCGGCTGCTGCGG 4140
 Db 14479 GCTCTGCCAACATCGCGGATGTGCG 14453

RESULT 4
 US-10-085-959-57/c
 ; Sequence 57, Application US/10085959
 ; Publication No. US20030165870A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Blatner, Frederick R.
 ; APPLICANT: Welch, Rodney A.
 ; APPLICANT: Burland, Valerie D.
 ; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
 ; FILE REFERENCE: 960296.97648
 ; CURRENT APPLICATION NUMBER: US/10/085.959
 ; CURRENT FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: 60/242,412

PRIOR FILING DATE: 2000-10-19
NUMBER OF SEQ ID NOS: 255
SOFTWARE: Patencin version 3.1
SEQ ID NO 57
LENGTH: 50538
TYPE: DNA
ORGANISM: Escherichia coli
US-10-085-955-57

Query Match 12.8%; Score 543.4; DB 12; Length 50538;
Best Local Similarity 48.5%; Pred. No. 1.1e-131;
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;

QY 181 ATCTGCCCCCCCCCGAGGCGGCGCTCCGTTTCTCTCAGACATCCAGAAATCC 240
DB 20201 ATATGTTCAGACCCCTGAGCATGTCATCAACCGTTCCATGTACGATTTACGCGGCC 20142
QY 241 TACTGGCTGGCGGACAGAGCGTTTACGTCGCCAGCGGATCCAGCCTATCGCAA 300
DB 20141 TACTGGCTTGAAG---ACAGAGCGGTGCTACTCTGATCGACCCATCTAACATGA 20085
QY 301 TACGACTGACGAGATCTCAACGTGCGGAGCTGAAGCGCGCTTTGCGAAAGTCGCG 360
DB 20084 TTTGACGTAGAACACTTATATGTTACGCGTTTACCATGCGGTGAATCGCTGATCGCT 20025
QY 361 CGGCGACAGATCTTCGGGCGCCACAGCGTGGCCGAGATGATGAGGTATCGACCTTAA 420
DB 20024 CGCCATGAATGTATGCTGCGGGGTATCTCCGAGGTACTAGAGATTCGCGCAA 19965
QY 421 GTGACGC---CGACATTCAGATCATCATGCTGCGGGCTCGACCGAGCACAGGAA 477
DB 19964 GTGCGGGGTATCATGTTAAGGAGCGCGATCTGAAGTCTTGTGCCCTTAACGAGAAAC 19905
QY 478 GCGAGGCTGTGTCTTGGGAGATGCGATGTGCGACCGCATCTATGACACCGAGCGCT 537
DB 19904 GATGCTTATGAGCGATTCGCGCATCGCTGTCGATCATGTGCATCCCGAGATCGTTGG 19845
QY 538 CGGCTTATCATGCTGTCCCGCTGGCTGAGACGAGCGCAACCGCTTCCGTCTCACT 597
DB 19844 CGGCTGTGTGATTTCACTTATTCGCTTGCATCGCGCAACATGCGCTTGCACTTTCAGT 19785
QY 598 ATGATCTCATTAACGTTGACCTTACGAGCAGCTGTCCATCATCTTCAAGAGACTGCTCAGC 657
DB 19784 CTGATCTGCTGATTTGCGATGCTCGATGATGCGACCGCTACAGAGAGATTGAGATG 19725
QY 658 TTCTACGAAGATCCGAGACTTCTCTCTGCTGCTGAGCTCTGTAACCGCATTAATGA 717
DB 19724 CTGTACCGTGAAGCCCATGCTGTCACTGCGTGTCTACCGTCTCTTTCGTGACTACGTG 19665
QY 718 CTCGCGCTGAGATCTGCAAGAAAGTGAAGGCGCATCAAGATCATGATGAATTAATGAAG 777
DB 19664 CAGGCGCTTTGTGAGAGAGGAGGAGTAAAGCTTATGACAGCATCAGGCTTATTTGCGAA 19605
QY 778 CGGCGCATCGCGAGCTCCCACTTCGCGCAGCGTTCGATGAAAGCGCATCATCTATACC 837
DB 19604 CGGCGCTGCGGAGCTGATGAGCCCAACCAAGCTCCCGTACAGGCGCATTTTGGCGAA 19545
QY 838 CTGAAGGAATCCGCTTCCGCGACAGAGCAATGCTGCGCTGCGACTTCCTGGGTGCA 897
DB 19544 CTGTCTGCATCAGGTTCTGTATCGTCCGCTCATCGCTGTCAAGCCACAACTGGGGAGTG 19485
QY 898 TTGAAGCGGCTGTGCGGGAGCGCGGCTGACCCCGAGCGGCTATCCGCGCTGATTT 957
DB 19484 CTGAGCGCTGTGCGCCACAGCACATATCAACAGAGCGCATTTGTTGACATGCTTT 19425
QY 958 TCCGAGGATCTCGGCGCTGAGACGCGAGCGCCCGGTTTACGCTCAACATAAGCTCTTC 1017
DB 19424 AGCGAAGTCTGCGAGTTGAGAGCTTAAAGCGGAGTTTACGCTCATGAGTTGTTTC 19365
QY 1018 AACCGGCTCCCGTCCATTCGCGCGTGAAGATATCAACGCGGAGCTTCAAGTCAATGTC 1077
DB 19364 AACCGGCTCCGAGGTTAACCCCAACGAGGCGAGTTATGTTGATTTACCGCTGTACAG 19305

QY 1078 CTCCTGACATGACACCACTCGGCGCAAGAGCTTGCAGACGCGGCTTAAGCTATTCAA 1137
DB 19304 TTGCTGAATGTTTGTGTAAGACAGCCGACTCTTATGCCCCACACGCTCAGGCTATTCCAG 19245
QY 1138 GAGCAGCTGTGGGAAGCCATGATCATCTGCGACGTAAAGCGTAT-----CGAGTTC 1188
DB 19244 GTGCAACTGTGGGAAGATCTGGAACATGCTGTTTCAATGGGATCGCGCCAGCGAGGGC 19185
QY 1189 CAGCGAAGCGCCCGGCTCTGGGATCCAGACGAGCGCATTTGTTCCCGTGTGCTC 1248
DB 19184 CTGATTCATGAGGCTGTCTGATTCAGTCCGCGATGCGGCTGATTAATGATTTGAT 19125
QY 1249 ACGAGCGGCTTAACAGAAAGTCTGTGTGTCACCTGTTGACAGGCTCGGAATCTCG 1308
DB 19124 ATGACGCGGAGAGACATCGCGCAACCTCGGAGACAAACCGTTTATCTGTGTCCG 19065
QY 1309 GTTACACACGACGACAGACTTCTCAGCTGCTGATATCATCAGCTTACGAGACAT 1368
DB 19064 GACGCCAATATTACCCAAACACCGAGGTGCTGATACCAAGTATGAGTTGGCT 19005
QY 1369 GGGGACTGCTGCTGCGGTGGACATCGTCAAGAGATGTTCCGCGGACTTGTGAC 1428
DB 19004 GGGGATTTGATTCATCACTGGGACGCGGTGCGAACTGTTGATTAACACGCTGCTGAT 18945
QY 1429 GACATGCTCGAAGCGTACGCTGTTTCTCCGCGGCTCACTAGAGAAACATGGGGTGA 1488
DB 18944 CAGATGTTGCTGCTTATTTGATCATGCTGAGGCGCTGTGGCCATGCTC-----G 18894
QY 1489 CAGTGCCTGTTGCTTCCGCTGCGCACTGAAGCGCGGCGAGCGCAACCGCAC 1548
DB 18893 CAAAGTGGTGGGGGTAAATGATTTCTGCGCGCTGCCAACCGTTAGTGCACGCTGACG 18834
QY 1549 AACGGCTGCTGAGGAGCATACGCTGACAGCGCTGTTGCGGCGCGGCTGAGACGCTG 1608
DB 18833 CAGGCTCTGCACTTACGCGCTTGTGCAACATGANTTATGCTGACGACACTGACG 18774
QY 1609 CCATGACGCTGCGGTGATGCTGCGCGCAAGAGCTTCACTGATGAGAGCTTTCGCGC 1668
DB 18773 CCACGAAACCTGCGCTGATCATGCTTATCGTGAATTTGACTTATGCGCACTGTGACG 18714
QY 1669 CGTTGCGGCGACTTGGCGCGCGGCTGCGAGAGGCGGCGACGCGCGAACATTTGTC 1728
DB 18713 GCGGCGGATCATGTGCGCGCGCTGTTAGGCTGCGGCTGACACATGCGACCGCGTG 18654
QY 1729 GCGGTGTGTATGAGAAAGCTGGAGCAGAGTGTTCGGGTTCTGCGGTGCTCAGATCA 1788
DB 18653 GCGGTGTGTATGAGAAAGCTGGAGCAGAGTGTTCGGGTTCAAGCGCATTTTACGACTG 18594
QY 1789 GCGCGGCTTACGTCGATCGATCGATCCGACCTTACCGGCGAGCGTATCCATCACTCTCTC 1848
DB 18593 GGTGCGGTCTATCTCCAGTGAATCCGATGCTACCGCACAGCGTCCGACGTTTGTGCTG 18534
QY 1849 GATCATGTGAGTAAAGCTGCTGTGACGACGCAATGCTGTGATGAGCAATCTGTATG 1908
DB 18533 ACGGTGGCGAGGTGCGGGTCAAGTAAACGCGCGGCTTCAACCAATTTGAGCGCTGCG 18474
QY 1909 CGGCGGGGATCCAGCGCTGCTGAGCGAGCGCGCGCTGAGAGGCGAGCGAGCAG 1968
DB 18473 CTGCGCG-----TGTGATCATGCGACGAGAAATCTGAGACAGCGCTCTGCG 18426
QY 1969 CTTCCGATGATCCCATTCAGACCTTTCGATCTCGGTATGATCATCACTTCGGA 2028
DB 18425 CCGTTCCTGAAGTGGCTGGGATGTCACAGATCTGGCTTATATCATTTTTCCTTCGCGC 18366
QY 2029 TCCACAGGTTGCCCAAGGGGTGATGATCATCTCGGGTGCCTCAACCATCTCTG 2088
DB 18365 TCCACCGTAAACCCGAAAGAGATGATGATGACACCGCGCGCATGAAACGCTGGA 18306
QY 2089 GACATCAAGAGCGCTTGCAGAAATAGGCGCGGAGCAGGCTGCGCTCTCTCGCTG 2148
DB 18305 GACATCAAGAGCGCTTGGCTCAATGCGAGATAGGATGTTGCGGCTGTCAATCATTTG 18246
QY 2149 ACCTTGATCTCTCGGTCTATGATGTGTTGCGGATCTGCGCGCGCGCGGTACATCGTG 2208

18245 AGCTTTGACCTGCTGGTTAGCATGCTTTGGCCCTTTTATGTTGGTGCAGCCGCTGTGTA 18186
OY GTGCGGAGCGGCTCAAGCTGGCGATCGGCGCATTTGGGCAAGATTGATGGAACGAG 2268
Db 18185 CTGCGGAGAGCAGCGGAAAAAGATCGCTCATTTGGCAGACAGTTATGACACAGCT 18126
OY 2269 AAGTGACGATGTGMACTCGGTGCGGCGCTGATGCGAATGCTGTGAGCATTTTGG 2328
Db 18125 CATGTAACTGTGAAATGCAAGTGCCTGACATGATGCAATGCTGTGG---AATACAC 18069
OY 2329 GGTGCCCCGATTGCTGCTAGTCTCTGGGCTTTGGCTGTGAGCGGCACTGGAATC 2388
Db 18068 AGCGGCGATCGGATGATTATCCGACGTGCGTGGCACTGTTGAGCGGCGACATGATC 18009
OY 2389 CCGGCGGCGCTGCTGCGGCA---GCTCCAGGCAATCAGGCCCGGCGTGTGATGAC 2445
Db 18008 CCGCTAACCTTACCGGAGCAGATGCGGACGCGCTCAATGAACGATGACATCATCAGT 17949
OY 2446 CTGGCGGCGGCGCACCGAAGCGTGCATCTGATCATCGGGTACCCCGTGGAGAACCTGCAC 2505
Db 17948 CTGGGTGGAAGCAGCCGATGCGCCATCTGTGCTGTACTACCGCATAGTGAAGTGA 17889
OY 2506 CTATGTGGGCGAGCATCTCCCTACGCGCGTCCGCTGCGCAACAGACGTTCCAGCTGTC 2565
Db 17888 TCGACGTGACCAATATCCCTACGCGTGGGCGCTGCGCAACAGCATATACGTGCTA 17829
OY 2566 GATGAGGCGCTGCMAACCGCGCCCGCTGCGGTTCGGGGGCAACTTAATTTGCGGGTTC 2625
Db 17828 AATGGCAACTGGAAGAAATGTCGGTTCGGGTGGAAGAGATTGATTTGGCGGAGT 17769
OY 2626 GGGCTGCACTGGCTACTGCGCGATGAGAAGAGAAGCGCAAGAGCTTCCCTGTCGAC 2685
Db 17768 GGGCTGGAACAAGCTTACTGAACGACGAGAAAGCGGCGCAAGCTTTGTCTGGGCG 17709
OY 2686 CCCGAGACCGGAGAGCGCTCTCAACAACCGGCGATCTGGGCGCTACCTGCGCATGA 2745
Db 17708 GAAAGGATGATGAGCGAATTTACCGCACTGGGAGATCGGCGGCGCTACTTTGCTGACGG 17649
OY 2746 AACTTCGATTCATGGGGCGTGAAGACAACCAATCAAGCTTCGGGATATCCGCTTGA 2805
Db 17648 CAAGTCGCTTTTGGGCGCAACGATCCCAAGGAGGTGAATGTTACCGTATCA 17589
OY 2806 CTGCGGGAATTCGAGAAACGCTCAAGTCGATCCGAAGTACGCGGATGTTG 2865
Db 17588 CTGGGGGAATTCGAGCGCTGATTCGCGCAATCCGATGTGAGCACTCACTGTGTG 17529
OY 2866 CCGCTCGGGAACGACGCGCGCAACAGCTCTTTCAGCTATGTGATCCGAGAGGACA 2925
Db 17528 GCAGTGGGTAA-----TTCTCAACATGTCGG 17502
OY 2926 CCGAGACCGCGCTGCGAGACAGACCGGAGCTCAAGACCGAGCGATGAGCGGAGACA 2985
Db 17501 CTGTGCTTTTTCGCAACTGACATGCGCACCGGCGCAGGCAATTCAGACTTAAGAA 17442
OY 2986 CACGCGCGCAAGCGGAGCTTGAAGCAGCGGAGGAGTGCAGTTCAAGTGTGCTGA 3045
Db 17441 -----GCGAGGCGGCGGCACTGCGAGGTATTTATGTGAATCCGCGACAGCTCTA 17388
OY 3046 CACGGAATTCGAGAGGACCTGACGGAAGCCGCTGCTGATCTGACCGGAGAGATCG 3105
Db 17387 GCGTTCAATCTCAAGAGACCAATATTCGCGCGCTGATGATGCTGGGATGCACTGACG 17328
OY 3106 CCGGAGGCGGCGCTGAGATCTACGCGCGTCCCTTACCTCCGAAGCTTCTTGAGGCC 3165
Db 17327 GCACCGGCGGATGACACAGTTACATCAACGCGCACCTATGCTATTTACGCGGCA 17268
OY 3166 CCGATTCCGTTTGTGATTTGATGATTTCTGAGCTGCTTGAAGAGCGTGAAGCCGAC 3225
Db 17267 AAAACCAAGCTGTGACAGATTTGGGCAATTTGCTGTGGGCTTGGGCAATGCTTACC 17208
OY 3226 GCGCGACCTTCCCAATTCGTTATCATGCGGCGGACAGCATACCGGTGCAAC 3285

Db 17207 GGTCTACCTTTTGGCAAGTATGCTATGCTCCGCGGCGGCGCTATACCCGTCMAAC 17148
OY 3286 TACGCTATGATCAATATCCGCGCATGAGGCGCTGACAGAGGCTTCTATTTATTCAC 3345
Db 17147 TACGTATCTGCTATCCAGACAGATGGAAGGAGATCCGATTTACTTCTGAC 17088
OY 3346 CCGTTGAGCACCGTTGCTGAACT---CTCCATCAAGGATTCAGCGGAGCGCAC 3402
Db 17087 CCGGACAGAGCTGCTTATGCTCCGATAGCACCAAGGCTGAGACAGTGTTCAT 17028
OY 3403 GTTGGCAAACTTCGACGCTTGTGATGAAGCGGCGTTCAACTCTCTGTTGGGACG 3462
Db 17027 GCCGACCTATACGCTATTTGCGGATCGGCGGCAATTCAGCTGTTATGCTGAT 16968
OY 3463 ATGACGCAATCGATGCTGTATG---ATGCTGTGCGAGAAATTTGCTGCTGAG 3519
Db 16967 ATGCGGTGATCTGCTCAATTTATGAGGAGAGAGGAGCTTGGCACTTCTCGTATGA 16908
OY 3520 GCCGATATATGCGGACGCTCTGATGAGAGCGGCGCTTCTGCAACATCGGCGTCTGT 3579
Db 16907 GCAGGTACTCTGCTCCATTTATCTGGAAGAGATGCGCGGCTACGATTTGGGCTGTG 16848
OY 3580 CCGGTGGCAATTCATTTTGAACAGTTGCGCGCTTCTGACCTGCACTTCGAC 3639
Db 16847 CAACTGGGATGCGACATTTTCCGCTGTGCAATTCGATTTCAATTTGCTGCCACATCAT 16788
OY 3640 GTTACGTCACCGCATGCTGGGCGGCGGCTAGACCCGCGGCACTTCAGTGTATG 3699
Db 16787 CGCTATGCTCATTCACCGTGGGCGCGCATAGGCGCAAGGCGGCAAGTGTGACCA 16728
OY 3700 CTGCGTGAAGATTCCTGACCGGAGCGGCGCACGACGCGGCGGCGCTCCGCGCGAG 3759
Db 16727 TTGCTGGCGCATTTCTCACTATGAGAAACCGAAGAAACCGCTGCGCGCTGAGCATG 16668
OY 3760 CAGCACTTCCCGCATTTATGCTTCCGCACTTTTGAAGCAAACTACCGGATACGTGTG 3819
Db 16667 CAGAGCTTACAAAGATGCTGCTGCGGCGCTGCTGACGAATCTGCTGATATATGTTG 16608
OY 3820 CCTACGCTTTCGAGAGCTGATGCTGCGGCTGCGTGCACCGGCAAGTGTGATGCT 3879
Db 16607 CCGAGTATCTGATGTTAGGACCACTTTCGTTAACCGTTAACCGGCACTGATGCTG 16548
OY 3880 AAGGCTGCGCGAGCGGAAAGATCTGTCGCGCGCAT-----TCGGGCGACAG 3933
Db 16547 CAATAATTAACGCTGAGGAGGAGAAATTTGCCACAGGTGACGCGGTGCTCAATC 16488
OY 3934 GCGCAAGGAGCGCTTGAAGAGATCTGTCGCGCTGCTGACGGAAGTCTGCGGCTG 3993
Db 16487 CAGGTGACAGTGGCTTACAAACAGCGGCTGTGCGCTTGGCAGAGTACTGCGGCTG 16428
OY 3994 GAGGTGCTGGGCTCCAGCAGAGCTTGTGATCTTGTGTGACATGATTCACATGTT 4053
Db 16427 AGCCAGTGTGCGCGAAGAGATTTCTTCTGCTGGGCGCAATTTATGAATGTGTG 16368
OY 4054 CGCATGAGAGCGCTGTTGCAAGAGGCTGATAGGAGATCGCATCACCGAGTCTTC 4113
Db 16367 CGTATTCAGAGGACCTTGAAGCGATTTATCGGAGAGATTTCCATTTGCTGATCTGTC 16308
OY 4114 CAGTACCGAACTTCGCGCTGCTGCGG 4140
Db 16307 GGTGCGCAACATCGCGATGTGGCG 16281

RESULT 5
US-10-085-959-54
; Sequence 54, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648

```

; CURRENT APPLICATION NUMBER: US/10/085,959
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/242,412
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: Patent version 3.1
; SEQ ID NO 54
; LENGTH: 32159
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1075)..(10795)
; OTHER INFORMATION: unsure
US-10-085-959-54

Query Match      9.2%; Score 390.4; DB 12; Length 32159;
Best Local Similarity 51.2%; Pred. No. 9.4e-92;
Matches 1086; Conservative 0; Mismatches 997; Indels 38; Gaps 6;

Qy      761 CGATGATTACTGGAAGCGCGCCGATCCGAGCTCCACCTCCGCGACCTTCCGATGA 820
Db      10021 CGCGGCTTACTGCTGGGGAAGCATCAGCGCTTCCCGCGCCGCTTGG 10080
Qy      821 AGCGGATCATCTACCTGGAAGAGATCGGCTCCGGACAAGAGCAATGGCTGCGT 880
Db      10081 CCTGGAAACCGCGACGCTATGTGAAGTCGTAATACCGAGCCGCAATGATTGTCGG 10140
Qy      881 CGGACTCTCTGGGATCGATTGAAGCGCGCTGTGCGGGAGCGCGGCTACCCGACGGGCG 940
Db      10141 CAACACGCTGGACGCGCTTTAGAACCGGGCGGAGATAGCGGTGACCGGACAAATG 10200
Qy      941 TCATCTGCTGCTGATTTTCCGAGGTATGCGGCGCTGGAGCGCGAGCCCGCTTACGC 1000
Db      10201 CGCTGGCGACCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10260
Qy      1001 TCACATACAGCTCTTCAACCGGCTCCCGTCATCCGCGCGTGAACGATATCACCGGG 1060
Db      10261 TTAAACATCACTTATTCAGCCGCGACGCTGCAACCGCGCGTGGCGCGGATGCTTGGCG 10320
Qy      1061 ACTTCACGTGATGCTCTCTGAGCATGACACCACTCGCGACAAAGACTTTCGAAACAC 1120
Db      10321 ACTTCACCAATATTTCTGCTGAGACACCGCTGCGATGCGATACGTCAGCAACCTGG 10380
Qy      1121 GCGCTAAGCTTTTCAAGACAGCTGTGGAGAGCATGATCACTGCGACGTAAGCGGTA 1180
Db      10381 CGCGTAA---AAACAGCTCAAGTTTACGAGGACTGGAGCATTCGCCACTGGTCGGCG 10437
Qy      1181 TCGAGGTCAGAGAGGCGCGCGGCTCTGCGGATCCACGAGGCGCATTTGTTCCCG 1240
Db      10438 TCGAATTACTCGGTGAATCAACGCGCAGAGCGCTAACCCCAAGGCGC-----CCCG 10491
Qy      1241 TGGTCTCAGAGCGGCTTAAACCAAGTCGTTGCTGTCATCTGTTGACAGAGCTCG 1300
Db      10492 TGGTATTTACAGCAATCTGGGGCGTTCCCTTACAGCAGCGCGCAGAAATGCGCGTGG 10551
Qy      1301 GAACTCCGCTGACACAGACGACATCTCTGAGTCTGCTGATCATCACTTACG 1360
Db      10552 GCGAGCGGAAATGGGCACTCTGCAACCGCGCAGGTCTGGAATCATCTGGCCCTTG 10611
Qy      1361 AGCAGATGGGGGCTGCTCTGCGGTGGGACATGTCGACGAGAGGTTCCCGCGCGAC 1420
Db      10612 AGCATATCGGCGAGTCTGCTGCTCAATGGAACGAAACAGACGCGCTGTTCTCCCGCGT 10671
Qy      1421 TTCTGACGACATGCTGAGAGCTACGTCGTTTCTCCGCGCGCTACACTGAGAACAT 1480
Db      10672 TAGTGAAACATGTTTGAAGCTTACGCAAGTTGATTAACCAACTCTGGAGAGAGAA 10731
Qy      1481 GGGGTGAACAGGTGC-----GCTGTTGCTTCCGCTCCGCACTGAGAGAGCGGGC 1532
Db      10732 GCGCCCTGGGAAAGCGTTCGAGATATTATGCCCGCCAGCCGCGGATACGGGAAC 10791
Qy      1533 GAGCGCAACGCGACCAAGCGCGTGTGAGAGGACATACGCTGACAGGCGCTGTTCCGGC 1592
```

```

Db      10792 GGAATCAACCCACCGGCTGCCCATTCGCCAAGGCTTGTGATGAAGCATTTTCCG 10851
Qy      1593 GCGGTGACAGACTGCCCATGACGCTGCGGTGTGCGCGCGCAAGAGCTTACGTA 1652
Db      10852 TATCGCTTCAACAGCCGACAGCGCTGCGGTAAACGACATGCTTATTCAGTGAATTA 10911
Qy      1653 CGAAGAGTTTTCGCGCGCTTGGCGCGGCGCATTTGGCGCGGCGGCTGCGCAGAGGGGCG 1712
Db      10912 TCATAGCTGACAGACTATGCTCCCGCTTGGCGCGGCGGCTTAATGAGTGGGGATTCA 10971
Qy      1713 CCCGAACATTTGTGCGGTGTGATGAGAGAAAGCTGGAGACAGTGTGTCGCGTTCT 1772
Db      10972 GCCCGGCAATATAGGTGCTATACAGATGTCGAAGAGCCAGAGACAACTTGTTCGCTTCT 11031
Qy      1773 CGCGGTCTCGAGTCAGCGCGGCTTACGTGCCATGATGCCACCTAACCGGCGGACG 1832
Db      11032 GCGCTCTGCTGCGCGGCGGCGGCTTACGTTTCGCTGATGATCAGCTGCGCGACG 11091
Qy      1833 TATCAGTACCTCTCGATCATGATGATGAAGTGAAGCTGCTGACGCAAGCATGCTGA 1892
Db      11092 GCGGAGAAATCTAGCTGAGCCAGCGCTGCGGTGCTCATTTGTGACAGACG 11151
Qy      1893 TGGCAAACTGTATGCGCGCGGATCCAGCGCTGCTGTGAGCGAGCGCGGCTGA 1952
Db      11152 CAGCGCGGCTCA-----GAGCATATTCGCGCTTGCCTGGCAGACAGGCCATT----- 11200
Qy      1953 AGCGACGCGACACAGCTCCGATGATGCCATTTAGACACTTGTGATCTGCGTATGT 2012
Db      11201 ---GAGCGGAGCGGATGCGCAACCCGCTAGTACGCGCCCGCCAGCAACCGGCTTACAT 11256
Qy      2013 CATCTACACTCGGATCCACAGAGGTGCGCAAGGGGGATGATGATCATCGGAGTGC 2072
Db      11257 TATCTACACTCTCGGCTTACCGGTACGCTCGAGGGGTAGTCACTTCTACCGGGAGC 11316
Qy      2073 CGTCAACACCATCTGACATCAACAGCGCTTCGAAATAGGCGCCGAGACAGAGTGTCT 2132
Db      11317 GCTTAACACGCTGTGCAATCAATACCGCGTATCAGGTGGCCGCAATGACAGGCTGCT 11376
Qy      2133 GCGCTCTCTCTGCTGAGCTTGCATCTCGGTCTATGATGTGTTGGGATCTTGGCGGC 2192
Db      11377 GCGCTCTCTCGCTTACATTTGATTTTACGACATTTTGGCGTACTGCGGC 11436
Qy      2193 GGGCGGTGAGTCGAGTGGCGGAGCGGTCACAGCTGCGCATTCGCGCATTTGGGACGA 2252
Db      11437 GGGCGCGGCTGAGTGTGATGATAAATACAGCGCGCATCTTACAGCATGTGTGA 11496
Qy      2253 GTTGATCGAACGAGAGAGTGAACGCTGAGTCCGCGCTGATGCGGATGCT 2312
Db      11497 GCTGATCCAGCGCATCAGTACAGTACGCTGGAACAGCGTCCGCGCTGTTGATATGCT 11556
Qy      2313 CGTCAACATTTTGAAGGTGCGCGCATTCGCTGCTGAGTCTCTGCGCTTCTGCTGCT 2372
Db      11557 GCTGACCTGTGTGAGAGTTTTCGCGACGCGCAAGCGGAAACCTGCGCGCATATGCT 11616
Qy      2373 GAGCGGCACTGGATCCCGGTGGGCTGCTGCGAGCTCCAGGCACTAGGCGCGGCT 2432
Db      11617 TTCGCGAGCTGATGCGGCTTGACTTCCCGCGCTTATGCGGCTTTCGCGCACAGG 11676
Qy      2433 GTTGTGATCACTGCGGCGGCGACCGAAGCGTGAATCTGCTGCTCATCGGTAACCCCT 2492
Db      11677 ACAATTTATCGGATGAGGCGGCGCACAGAGCGTATCTGCTGTAACGCTCGGAAT 11736
Qy      2493 GAGGAACGTGACCTATCTGCGGCAAGCATCTCCCTACGCGCGCTCGCTGCGCAACAGC 2552
Db      11737 TCACGACGCTCCCGCCACTGCGCTCATCTTACGCTTTCGCTTAAACCAACCAAG 11796
Qy      2553 GTTCCAGCTGCTGATGAGGCGCTGAAACCGCGCGGCTGAGTTCCGCGGCAACTTA 2612
Db      11797 CTACCGGAGTGTGAGTAAGACAGGCGCGGACCTGCTTACTGAGTGGGTCGCGGTAATATG 11856
Qy      2613 CATTGGCGGAGTGGGCTGGACTGGGCTTACTGGCGGATGAAGAAACCGCGCAAGAG 2672
```

Db 11857 GATTGGCGCGCATTTGGGGTGGCGGAAGCGTATTATTCACAAGATCCCTCGCTACGACGACGA 11916

Qy 2673 CTTCCTCGTGCACCCCGAGACCGGGGAGCGCCTCTCAAGAAGCCGGCGATCTGGGCGCGCTA 2732

Db 11917 ATTTTTCGCTCCCGGA-----CGAAGCTGGTATGCGACACGGCGCATCTCGGCTGCTA 11970

Qy 2733 CCTCCCGGATGGAAACATCGAGTTCTAGGGGCGTGAAGGACAAACAAATCAAGCTTCGCG 2792

Db 11971 CTGGCCGAGATGGCACATCGAGTTCCTGTGCTGCGACGAAGCAGCTCAAGTAGTGGAGG 12030

Qy 2793 ATACCGCGTTAGCTCGGGGAAATCGAAGAAACGCTCAAGTCGATCCGAACGTACGCGA 2852

Db 12031 ATATCGCATCGAGCTGGGCGAAATCGAAAGCGCGTCAAGCAGCTGGGGGGTGAAGAA 12090

Qy 2853 CGCGGATATTGGCCCGCTCG 2873

Db 12091 AGCAACCGTTCTGGCGATCGG 12111

```

RESULT 6
US-10-314-657-1
; Sequence 1, Application US/10314657
; Publication No. US20030175888A1
; GENERAL INFORMATION:
; APPLICANT: SHEN, Ben
; APPLICANT: CHENG, Yi-Qiang
; APPLICANT: TANG, Gong-Li
; TITLE OF INVENTION: Discrete Acyltransferases Associated with Type I Polyketide
; TITLE OF INVENTION: Syntheses and Methods of Use
; FILE REFERENCE: 054030-0021
; CURRENT APPLICATION NUMBER: US/10/314,657
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: PCT/US02/08937
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: US 60/278,935
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 135638
; TYPE: DNA
; ORGANISM: Streptomyces atroolivaceus
US-10-314-657-1

```

Query Match	8.5%;	Score 360.6;	DB 12;	Length 135638;
Best Local Similarity	55.48;	Pred. No. 7,6e-84;		
Matches	768;	Conservative	0;	Mismatches 604;
			Indels	15;
			Gaps	3;
QY	1541	ACGCGACCAACGCGCGCTGCTAGCGAGCATATCGCTGACAGGAGCTGTTTCGCGCGCGGGATCG	1600	
Db	72583	ACACCAACCGCGCGCTTCCCGACGACCGGCTGATGCAACAGCTCTTCAGAGACAGCGCG	72642	
QY	1601	AGCAGCTGCCCATGACAGCTGCGCGGTGTGCG--GCGCGCAAGAGCTTACGTAAGAAG	1657	
Db	72643	CCCGACGAGCTCGCGGCGGAGAGGCGCTGCGCTGCGCGGCGCGGCGACAGATGACATACAG	72702	
QY	1658	AGCTTTCGCGCGCGCTGCGCGGCGACTTGGCGCGCGGCTTGGCGGACAGGGGGACAGCGCCA	1717	
Db	72703	AGCTGAACCGACCGGCGCAACCGCATGCGCGCGCGGCTGGCGCGGAGACGTGGGCGCTCG	72762	
QY	1718	ACAATTGATGTCGCGGATGATGAGAAAGGCTGGGAGCAGGTGTTCGCGGTTCTCGCGG	1777	
Db	72763	AGACGGTCGTTCGCGCTCAGTGTCCGCGCGGCGCCATGATAGTTCGCGGTGTCTTCGGGA	72822	
QY	1778	TGCTCGAGTACGGCGCGGCTTACGTGCGATCGATGCGGACCTTACCGGCGGAGCGTATTC	1837	
Db	72823	TCCGGAAGCGCGCGGCGGTGTCTATCTGCGATGAGCGCGCATCTGCGCGCGGAGCGGCGCG	72882	
QY	1838	ACTACCTTCCTGATCATGTGAGGTAAAGTTCGAGCGACGACGACATAGGCTGATGGCA	1897	
Db	72883	CCGTCATCTTCGAGAGAGCGCACCGCAGAGTGTGTCTTCATCACCGCGGACCGGAGGGCTT	72942	
QY	1898	AAC---TGTGATGGCGCGCGGAGTCCAGCGGCTGCTGTGACGAGGCGCGCGGTGGAAG	1954	

Db	72943	GGCCGGTGTCCGACGGATACGCGCCCGGTGTGGCGGAGCCGCGCTGTGAGGGCCCCCAC	73002
Qy	1955	GGCAGCGCGACACGACCTCCGATGATGCTCCGATTCAGACACTTCGATCTCCGATATGCA	2014
Db	73003	CCGCGCAGCGGAGCAACTGCGCCGCGCGCTGTGACGAGCCGACAAACACGGGATCATCA	73062
Qy	2015	TCTAACCTTGGGATCCACAGGGTTGGCCCAAGGGGGTATGATTCGATATCGGGGTCCG	2074
Db	73063	TCTTCACTCTGGGGCAGACACCGGCGGCCCAAGGGGTGTGGCGGTGCGCACCGGCGCTTC	73122
Qy	2075	TCAACCAACATCTGGGACATCAACGACGGTTGGAAATAGGGGCCGAGAGCAGGGTCTGG	2134
Db	73123	TGAACCTCAACAACTGGATGCGGCGGACCTTCGGGTTCCGGGCCCGGAGACATAGGCGCTGT	73182
Qy	2135	CGGCTCTCTGCTGAGACTTCGATCTCTCCGATCTATGATGTGTTCGGGATCTTGGCGCGCG	2194
Db	73183	GTGTGACCTGCTCGGCTTCGACCTGTCCGTGTGTGACGTGTTCGGCCGTCTCGGTACGG	73242
Qy	2195	GCGGTACGATCGTGTGGCCGAGCGGTCCAAAGTGTGCGCATTCGGCGCATTTGGGCAAGT	2254
Db	73243	GTGCGCGGCTCTCAATCGCGGACCGCGGACACAGCGCATCCGCTCTCTCTGTGACG	73302
Qy	2255	TGATTCGAACGAGAGAGTACGATGCGGTGGAACTCCGGTGCGGGCGGTGATGGGATGTGCG	2314
Db	73303	TGCTGATCGAGGAGCGGATCACTTTGGAACTCGGCGCCACACAGCTCGCCCAAGTACG	73362
Qy	2315	TCGAGCAATTTGAGGGTTCGCCCGATTCGCTGATAGTCTCTGCGGCTTTGCGTGTGA	2374
Db	73363	GTCCGGTGTCTGACACAGTGGGACGGGGGATACGGGTATCTGGGCGCTGTGTTCTCA	73422
Qy	2375	GCGCGCATGTGAATCCCGGTGGGCTGTGGCGCTGTGGGAGCTCCAGGCGATCAGGCCCGGTGT	2434
Db	73423	GCGCGCATTCAGCGCGCTGTGCCCTTCCCGAGAGGTGTGGCGCGCTTCCGCGCGCGCG	73482
Qy	2435	CGGATTCAGGCTTGGGGGGGGCCACCGAAGGCTCATCTGTGTCAATTCGGGTACCCCGTGA	2494
Db	73483	ACATGATACAGCTGGGCGGGGCGCACCGAGGCGACGGTGTGTCAACTGTGTTCGCAATCG	73542
Qy	2495	GGAACTGCACTTATGTGGGCGAGCATCCCTACCGCGGCTCGGCTGCGACCAAGAGT	2554
Db	73543	GGGCGATCGACCGGGCTGTGGCGGAGATCCCTTACGCGAGGCGCATCGACACACCGCT	73602
Qy	2555	TCCAGCTGTCTGATGAGGCGCTCGAACCGCGCCCGGTCTGGGTTCCGGGGACACTTCA	2614
Db	73603	ACCACGTGTCTGACGAGGCGGTGCGCCCTGTCTCGTCCGGCTCGAGGGGGATCTGTCA	73662
Qy	2615	TTTGGCGGGGTCTGGGCTGTGGCACTGGGCTAATGGCGGATGAAAGAAAGACGCCAAGACT	2674
Db	73663	TGCGCGGTGATGTCTGGGCTTGTGGCTTGTGGTATGTCACACGCGGAGCTGACCGCGGACGGT	73722
Qy	2675	TTCTCT-----GTGCACTCCGAGACCGGGGAGGCGCTTCAAGAACCGGCGATCTGG	2725
Db	73723	TCAATCCCGACCGGCTTCCADAGAGACCTCAAGAGGCGCTTCAAGAGACGGGTACAGGG	73782
Qy	2726	GCGCGTACTGTGCCCGATGGAAACATCGATTCATGGGCGTGAAGCAACCAATCAAGC	2785
Db	73783	CCCTGTACTACCGGACCGGCAACCTGAGCTTCAAGGGGCGGGCCGACGGAAGTGAAGG	73842
Qy	2786	TTTGGCGGATACCGCGCTTGTGAGCTTCGGGGAAATGAGAGAAACGTTCAAGTCCGATCCGAACG	2845
Db	73843	TTCGCGGGTTCGCGGTGGAACTGTGGCGGAATGAGACACGGCTGTGGGCGACGAGGGG	73902
Qy	2846	TACGCGACGGGTGATTTGTGCCGTCTGGGAAACGACGCGCGCAACAGCTCTTCTAGCT	2905
Db	73903	TCAAGGACGCGCGTGTCTGTGGCCCGGAGAGAGCGGTGTGGGACCGCAACCTCTGTGGGT	73962
Qy	2906	ATGTGT 2912	
Db	73963	ACCTGT 73969	

RESULT 7

APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 PRIOR FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 3143
 LENGTH: 1875
 TYPE: DNA
 ORGANISM: Streptomyces avermitilis
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)..(1875)
 US-10-156-761-3143

Query Match 7.0%; Score 296.8; DB 14; Length 1875;
 Best Local Similarity 53.0%; Pred. No. 1,66-67;
 Matches 743; Conservative 0; Mismatches 632; Indels 27; Gaps 4;

1533 GAGCGCAACCGGACCAACGCGCTGCTGAGCGAGCATACCTTGACGGCTTTGCGGCGC 1592
 72 GAACGACACCGCGGGAGCTGCGCGGCGGTGGGACACCTGCGCCGAGCTGTCCGTAC 131
 1593 GCGGCTGAGAGAGTCCCATGACAGCTCCGCTGTGTGCGCGCGGCAAGAGCTCAGCTA 1652
 132 GCAGGTGTCCCGAACCCCTGACGCCACCGCCCTGTCTTCGCGGACACCGCCCTACGTA 191
 1553 CGAAGAGCTTTGCGCGCTTGGCGGCGCACTTGCGCGCGCTGCGCGAGCGAGGGGACG 1712
 192 CACCGAGCTGACGAGCGCTGTGAGACCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGC 251
 1713 CCGGACACATTTGCTCGCGGTGTGTGATGAGAGAAAGCTGGAGAGAGTTGTGCGGTTCT 1772
 252 CCGGACCGGGGTCTGCGCGCTGACCGGTGCGCGCTGCGGTGAACTCGTGTGCGGCTGCT 311
 1773 CCGGCTGTGAGTACGCGCGCGCGCTACGTGCGCATGATGATGATGATGATGATGATGAT 1832
 312 CCGGCTGTGAGAGAGCGCGCGCGCTTACCTCCGCTGAGACCGGACCTACCGCGGAGCG 371
 1833 TATCCATCACTCTCTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1892
 372 CATCGCTACCTCTCTCGCGAGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 431
 1893 TGGCAACTGT-----CATGCGCGCGCGGAGATCCAGCGGCTGCTGAGCGAGGC 1943
 432 GCGGACCTGCG 491
 1944 CCGGCTGAGAGCGACG 2003
 492 GGAAGACCGCGTGAAGCG 551
 2004 CCGCTATGTCATCTACACCTCGGAGATCCACGAGGTTGCCAAGGCGGATGATGATGATGAT 2063
 552 CCGCTATGTCATCTACACCTCGGAGATCCACGAGGTTGCCAAGGCGGATGATGATGATGAT 611
 2064 TCGGGGTGCGCTCAACACCATCTGAGCATCAACAGCGCTTCGAAATAGGCGCGCGAGA 2123
 612 CGCGGATAGTCAACCGCGCTGCACTGATGATGATGATGATGATGATGATGATGATGATGAT 671
 2124 CAGGCTGTGCGCGCTCTCTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 2183
 672 CCGGCTGTGAGAGAGCG 731
 2184 CCGGCTGTGAGAGAGCG 2243
 732 GCTGATACCG 791
 2244 TTGGGCAAGAGTTGATCGAAG 2303

Db 792 CTGGCCGAGCTGATCCGGGCGAGCGGATCACCAACCGTCACTTCTCCGCTGATGCT 851
 Oy 2304 GCGGATCTGCTGAGAGATTTTGAAGGATGCGCCCGATTTGCTGCTGATGCTGCGGCT 2363
 Db 852 CTAGGTTC-----CTGAGCAGCGCGCGCGCGCGCGCGCTGTGACCGGCTGCGCG 902
 Oy 2364 TTGCTGCTGAGCGGCGATGATTCGCGTGGGCTGCTGCGAGCTGCAAGGCTCATGAG 2423
 Db 903 GGTGTGTGAGCGGATGAGCGGCTGCGCGTGTGAACTGCGCGACGCTTACAGGCGTGC 962
 Oy 2424 GCGCGGCTGTGCTGATTCAGGCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2483
 Db 963 GCGCGGCTGTGCTGATTCAGGCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1019
 Oy 2484 GTACCCCGTGAAGAGATGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2543
 Db 1020 GTACTGGAAGTACACCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1079
 Oy 2544 CAACGACAGTTCCACGCTGCTGATGAGAGCGCTGAAACCGCGCGCGCGCTGCGGTTCCGG 2603
 Db 1080 GAACACCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1139
 Oy 2604 GCACTTATGATGAGCGGCTGCGGCTGCGCACTGCGCTTACTGCGCGCATGAAGAGAC 2663
 Db 1140 CAGCTGTATCTGCGCGGCTGAGCTGCGCTGCGCTGATCTGAAACCGCGCGCGCTGAC 1199
 Oy 2664 GCGCAAGAGTTCTGCTGCTGATGACCCCGA-----GACCGGAGAGCGCTTACAAAGACCG 2717
 Db 1200 GCGGAAAGCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259
 Oy 2718 CGATCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2777
 Db 1260 GAGACTGCG 1319
 Oy 2778 AATCACTTTCGCGGATACCGGCTGAGCTGCGGGAATGAGAGAAAGCTCAAGTGC 2837
 Db 1320 GGTCAAGATTCGCGGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCG 1379
 Oy 2838 TCGGAAGTACGAGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2897
 Db 1380 TCGCGCGCTTTCGCGGCT 1439
 Oy 2898 TCTAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2919
 Db 1440 GTGGGCTATGTGTGCTCCGCGC 1461

RESULT 9
 US-10-156-761-1/c
 Sequence 1, Application US/10156761
 Publication No. US20030119018A1
 GENERAL INFORMATION:
 APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 PRIOR FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 1
 LENGTH: 9025608
 TYPE: DNA
 ORGANISM: Streptomyces avermitilis
 FEATURE:

NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 53.0%; Pred. No. 7.8e-67;
Matches 743; Conservative 0; Mismatches 632; Indels 27; Gaps 4;

```
QY 1533 GAGCGCAACGCGACCAACGCGCTGTGAGCGAGCATAGCTGCAAGCGCTTTCGCGC 1592
DB 3931113 GAACGACACCGCGCGGAGCTCGCGCGGTGGCGACACCTTGCCGAGCTGTCGTAC 3932054
QY 1593 GGGGCTGAGAGAGCTGCCATGACAGTCCGCTGTGTGTGGGGCGGCAAGACCTCAAGTA 1652
DB 3932053 GAGGCTGTCCGAGACCCCTGAGCGCAACCGCTGTGTTCGAGGACCCGCTCAAGCTA 3931994
QY 1653 CGAAGAGCTTTTCCGCGCTTTCGCGCGCATTTGAGCGCGCGCTGCGGAGCAAGGAGCAG 1712
DB 3931993 CACCGAGCTGAGAGAGCGCTTCGAGACCGCTTGCGCCGCGCTGTGGCCGAGCGGCGC 3931934
QY 1713 CCGGACACATTGTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1772
DB 3931933 CCGGACACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931874
QY 1773 CCGGCTGCTGAGTCAAGCGCGCGCGCTTACGTCGAGTGTGTGTGTGTGTGTGTGTGTGT 1832
DB 3931873 CCGGCTGCTGAGTCAAGCGCGCGCGCTTACGTCGAGTGTGTGTGTGTGTGTGTGTGT 3931814
QY 1833 TATCCAGTACTCTCTGATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1892
DB 3931813 CATCGGCTACTCTCTGCGGAGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931754
QY 1893 TGGCAAACTGT-----CATGCGCGCGCGGAGTTCAGCGCGCTGTGTGTGTGTGTGT 1943
DB 3931753 GCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931694
QY 1944 CGGCGCTGAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931634
DB 3931693 GGAAGAGCGCGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931634
QY 2004 CCGGATGTATCTATCACTCTCGGAGTTCACAGGCTTGTCCCAAGGAGGTGTGTGTGTGT 2063
DB 3931633 CCGGATGTATCTATCACTCTCGGAGTTCACAGGCTTGTCCCAAGGAGGTGTGTGTGT 3931574
QY 2064 TCGGGGTGCTGTCAACCATCTGTGACATCAAGCGCGCTTGTGAATAGGCGCGCGGAG 2123
DB 3931573 CCGCGGATATGTCACCGGCTGACATGTGACAGCGAGTACCGGCTGACCGCGGAG 3931514
QY 2124 CAGGGGTGCGGGCTCTCTCGCTGAGCTTGTGATCTCTGCGGTGTGTGTGTGTGTGTGT 2183
DB 3931513 CCGGCTGCTGAGAGAGACCGCTCTCGGCTTTCAGCGTGTGTGTGTGTGTGTGTGT 3931454
QY 2184 CCGGCGCGCGCGCGGTGACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2243
DB 3931453 GGTGATTCACCGGCGCGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931394
QY 2244 TTGGGCGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2303
DB 3931393 CCGTGGCGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931334
QY 2304 GGGGATGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2363
DB 3931333 CAGGCTGTCT-----CTGAGCGAGCGCGCGCGCGCGCTGTGTGTGTGTGTGTGT 3931283
QY 2364 TTTCGCTGTGAGCGCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2423
DB 3931282 GGTGATGTGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931223
QY 2424 GCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2483
DB 3931222 GCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931166
```

```
QY 2484 GATCCCGGTGAGGAACGTGTGACCTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2543
DB 3931165 GTACTGGAGTATACACCCCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3931106
QY 2544 CAACCGAGCGTGTTCACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2603
DB 3931105 GAACACCGCGCTGTATAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3931046
QY 2604 GCACTGTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2663
DB 3931045 CGAGCTGTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930986
QY 2664 GCGCAAGAGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2717
DB 3930985 GCGGAGAGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930926
QY 2718 CGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2777
DB 3930925 GGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930866
QY 2778 AATCAAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2837
DB 3930865 GGTCAAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930806
QY 2838 TCCGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2897
DB 3930805 TCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930746
QY 2898 TCTAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2919
DB 3930745 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3930724
```

RESULT 10
US-10-156-761-3184
Sequence 3184, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 3184
LENGTH: 5304
TYPER: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1) .. (5304)
US-10-156-761-3184

Query Match
Best Local Similarity 6.4%; Score 270.4; DB 14; Length 5304;
Matches 753; Conservative 0; Mismatches 656; Indels 39; Gaps 5;

```
QY 1503 GTTTCGCTGTGAGCGCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1562
DB 741 GCGAGCGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800
QY 1563 CGAGCATACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1619
```

Db 801 GGACACCGTGGCCCTCGACGAGCTTTTACCCCGGAGGGGGGCTGAGCCCGGAGCCAC 860
Qy 1620 CGCCCTGTGTGTGGCGCGCAAGACCTGACGTAAGAGCTTTTGGCCGCTTGGCGCG 1679
Db 861 CGCCGTGTGTGGCGCGCAAGACCTTCCGGCTTACCTACCCGAACTCGAGCCGCGTCCAA 920
Qy 1680 ACTTGGCGCGCGGCTGGCGGAGCAGGGGACCGCCGAAACATTTGGTGGCGGTGTAT 1739
Db 921 GCTGGCGGAGCGCTGGCGGAGCGGGGGCGGCGGAGCACTGTGGCGGTGTAT 980
Qy 1740 GGAAGAAAGCTGGGAGGAGCTTGTGGCGGTCTTCCGGGTCTGAGTGAAGGCGCGCTTA 1799
Db 981 CGACCGCTGGCGCGAGCTGTGTGGCGGTCTTACCGCTTCAAGAGCGCGGCGCGCTTA 1040
Qy 1800 CGTGGCGAGTGGCGCGAGCTTACCGCGGAGGAGCTTACCTACCTCTGATGATGTGA 1859
Db 1041 TCTGGCGGTGAGACCCCGGCTACCGCTGAGGAGCTTCCGCTACGCTCTGGAGAGAGCG 1100
Qy 1860 GGTAAAGCTGTGTGAGCGAGCGAGCTGTGATGGCAAACTGTATGGCGCGCGGGAT 1919
Db 1101 CGCTGTCTGTCTGAGCGAGCGCGCGCGCGCGCGCGAGCTT-----GATTCAGAGAG 1151
Qy 1920 CCAAGCGGTGTCTGTGAGCGAGCGCGCGGTGAAAGCGAGCGAGCGAGCTTCCGATAT 1979
Db 1152 CGCGAGGTACTGTGAGCTGAGCGAGAGTCCGCTGACCAAGAGAGCGAGCGCGCGA 1211
Qy 1980 GCCCTTACAGACACTTTCGATCTCGCGTATGTCTACACCTCGGGATCCAGAGGTT 2039
Db 1212 GGTATCGGCGAGCGCGCGCGAGCTTGGCTGATCTACACTCTGGGCTGAGCGCGA 1271
Qy 2040 GCCCAAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2099
Db 1272 TCCCAAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1331
Qy 2100 GCGCTTCAAAATAGGGCGCGGAGAGAGAGGTGTGGCGCTTCTCTGCGTGAAGTGTAT 2159
Db 1332 CGCTTACCGCGTGTGAGCGAGCTGAGTATGAGAGAGAGCGCGCGCTTCTGAGAGT 1391
Qy 2160 CTCCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2219
Db 1392 GTCCGTCTGGAGAGCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1451
Qy 2220 GTCCAAAGTGTGCGAGTCCGCGCGCTTGGCGAGAGTGTGAAAGAGAGAGAGTGTAT 2279
Db 1452 CGGCGCGAGAGCGCGAGCGCGAGCGCGGTGTACCGCGAGCGGTGACGAGTGTAT 1511
Qy 2280 GTGGAAGCTGT 2339
Db 1512 CATGACATCTGT 1571
Qy 2340 TTCCGTCTGAGAGTCTGTGGCGCTTTCGCTGTGAGCGCGAGTGTATCCGGTGGGCT 2399
Db 1572 ACAAGCGCGGT 1631
Qy 2400 GCTGGCGAGTTCAG-----CATCAGCGCGCGGTGTGTGTGTAT 2441
Db 1632 CCAGCTCGCGCGCTTCCGCGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1691
Qy 2442 CAGCTGTGGCGCGCGCGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2501
Db 1692 CAACCTGTAGCGCGCGCGAGCGAGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1751
Qy 2502 CGACCTATGTGTGGCGAGATCCCTTACGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2561
Db 1752 CGAGCT-----GAGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1805
Qy 2562 GCTGTAGAGCGGT 2621
Db 1806 CCTGT 1865
Qy 2632 GGT 2681
Db 1866 TGT 1925

Qy 2682 GCACCCGAGACCGS---GAGCGCTTCTACAGACCGCGATGTGGCGCTACTGCGC 2738
Db 1926 CGCCCGCGCGCGGTGTGAGAGCGCGTCTTACGAGCGGCGAGCTGGCGGTGGTGGC 1985
Qy 2739 CGATGAAACATCGAGTTTATGTGGCGGTGTGAGAGCAACCAATTAAGCTTCCGATACG 2798
Db 1986 CGAGCGGAGCATTCACCTTGTGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2045
Qy 2799 CGTTGAGCTGGGGAAATCGAGAAAGCGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGT 2858
Db 2046 CATGAGCTGTGGCGAGATGTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2105
Qy 2859 GATTGTGCGCTGTGGAGAGAGCGCGGAGCAAGCTTCTTACCTTATGTGTGTGTGTGT 2918
Db 2106 GT 2165
Qy 2919 GGGGACAC 2926
Db 2166 GGACGCGC 2173

RESULT 11
US-09-815-242-4052
Sequence 4052, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4052
LENGTH: 13029
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4052

Query Match 6.2%; Score 261.4; DB 9; Length 13029;
Best Local Similarity 51.8%; Pred. No. 4.1e-58;
Matches 704; Conservative 0; Mismatches 631; Indels 24; Gaps 4;
Qy 1572 GCTGCAAGCGCTGT 1631
Db 3387 GCTGCGGAGGT 3446
Qy 1632 GCGCGCAAGCGCTGT 1691
Db 3447 GAGCGGCGAGCGCTGT 3506

QY 1692 GCTGCGGAGCAGGCGGCAAGCCCGAACAATTGTCGGGTGATGAGAAAGCTG 1751
 DB 3507 CCTGGCGCGAAGAGGGGCTGGGCGCGGACCTGGGGGAGGATCTGGCGCGGTTGCGC 3566
 QY 1752 GAGAGAGGTTGTCGGGTTCTGGCGGTGTCGATGAGGCGGGGCTTACGTGCCATGCA 1811
 DB 3567 GCACTGCTGTGGTCTGTGCGGATCGTCAAGGCGGGGCGGTGCTTACGTGCGCTGGA 3626
 QY 1812 TCGCGACTTACCGCGGAGCGTATCCACTTACCTCTCGATCAATGATGAGTAAAGCTCGT 1871
 DB 3627 CCGGAGTACCCGAGGAAAGCGCTGGCTACATGCTCGCGACAGTGGGCTGAGAGTGT 3686
 QY 1872 GCTGACGAGCCATGCTGATGAGCAACTGTGATGAGCGCGGGGATCCAGCGCTGCT 1931
 DB 3687 ACTGACCCGAGCCCATCTGTTCGACGCGCTGCGCGGAGGGGCTGACGCGCATCTG 3746
 QY 1932 CGTGAAGAGGCGCGGCTGGAAGGCGAGCGGCGACGCGCTCGATGATGCCATTGAGAC 1991
 DB 3747 CCTGAGACGCTTACGACTGACAACTGGCGGAGCCGAGCCGAGCGGCGCTGACCTGACG 3806
 QY 1992 ACCTTGCATCTCGCTATGATCTTACACTTCGCGGATCCAGAGGTTGCCAAGGGGCT 2051
 DB 3807 CGACA---ATCTGCGCTAGTATCACTCACTCGGCTCCAGCGGCGACCGAAGGGCGCT 3863
 QY 2052 GATGATGATCATCGGGGTGCGCTCAACACCATCTGGAATCAAGAGCGCTTGAAT 2111
 DB 3864 GGGCAACACCCACGCGCGCTGCGCGCTGCGCATGAGATGCAAGCCCATCAACCTCT 3923
 QY 2112 AGGCGCGGAGACAGGGGTGCTGCGCGCTCTCTCGGAGCTTCGATCTCTCGGTCATG 2171
 DB 3924 CAGCGGCGAGAGTCTGATGAGAAAGGCGCGGAGCTTCGATGCTGCTGAGG 3983
 QY 2172 TGTGTCGGGATCTGCGCGCGCGCGGCGATGATGATGATGCGCGAGCGCTCAAGCTGCG 2231
 DB 3984 GTGCTTCTGCGCGCTGTCACGCGTTGCGGTTGCTGCTGCGCGCGCGCGGAGCACCG 4043
 QY 2232 CGATCCGCGGATTTGGGCGAGAGTTGATCGAACGAGAAAGTACGCTGTGAATCGCT 2291
 DB 4044 CGATCCGCGGCGCTGCTGGAACCTGTGCGCGCATGTCGAGGCTGACCTGACCTTCT 4103
 QY 2292 GCGCGGCGTGAAGCGGATGCTGTCGAGCAATTTGAGGGTGGCGCGGATTCGCTGCTAG 2351
 DB 4104 GCGCGCGCTGCTGCACTGTTATC-----GACGAACCGCGGCTGCGCGCTGCGG 4154
 QY 2352 GTCTCTCGGCTTTGCTGCTGAGCGGCGAGCTGATCCGCGTGGCTGCTGCGGAGCT 2411
 DB 4155 CAGCTTGGCGCGCTGTTCTCCGCGCGGCGAGGCGCTTCGCGGAGACTGCGCAACCGCT 4214
 QY 2412 CCAAGCGCATGAGCGCGCGGTGCTGATCAAGCTGGGCGGGGCGACCGAAGCTGAT 2471
 DB 4215 GTTGCAACGCTGCGCGGCTGCTGATTAACCTTACGAGCCCAACCGAAGCCGCCAT 4274
 QY 2472 CTGCTCATGCGGATCCCGGTGAGAAAGTGAAGCTATGCTGGGCGAGCATCCCTTACG 2531
 DB 4275 -----CAAGCTTACCATTTGGCAGTGGCGCGGAGAGCGGAGCGCTTCCCTGATCG 4328
 QY 2532 CCGTCCGCTGCGCAACAGAGCTTTCAGAGTGTGATGAGAGCGCTGAAACCGCGCGCT 2591
 DB 4329 CGAGCGGCTGAGGAGAGTGTGCTGCGCGGTGCTGAGAGCGGATTAACCTGCTGCGG 4388
 QY 2592 CTGCGGTTCCGGGCAACTTACATTTGGCGGGGTGCGGCTGAGCACTGGGCTTACGCGCA 2651
 DB 4389 CGGCGTGCAGGAGAGCTGTGATCGCGGCTGAGGCGCGGCTGAGCACTTGGGCGCG 4448
 QY 2652 TGAAGAGAGAGCGGAGAGCTTCTCG-----TGCAACCCGAGAGCGGGAGGCGCT 2705
 DB 4449 CCGGCGCTGAGGCGGAGCGCTTCTGCTGCGCATCCGTTCTGCGCGCGCGGAGAGCGCT 4508
 QY 2706 CTACAAAGACCGGAGCTTGGGCGCTACCTGCGGATGAGAAACATGAGTTCAATGGGCG 2765
 DB 4509 GTACCGCACCGGAGCGGCGGCTGAGAACGCGGAGCGGCTGAGTAATCTCGGCGG 4568

QY 2766 TGAAGACAACCAATCAAGCTTCCGCGATACCGGCTTGAAGCTCGGAGAAATCGAGAAAC 2825
 DB 4569 CTTCAACAGAGCTCAAGCTGCGGTCTCCGATCAAGCGGAGAGAAATCAAGGACG 4628
 QY 2826 GCTCAAGTGCATCCGAAGTACGCGAGCGGATGTTGTGCCCTCGGAAAGAGCGCGC 2885
 DB 4629 CTTGTGCGGAGAGCGGGGGGTGGCCAGGCGGTAGTGTGATCCGCGAAGCGTGGCGG 4688
 QY 2886 GACCAAGCTCTTCTAGCTTATGTGATCCGAGGCGAC 2924
 DB 4689 CAGCGATTTGTGCGCTACTACCGGCGCTGTGCGCGC 4727
 RESULT 12
 US-10-156-761-3629
 / Sequence 3629, Application US/10156761
 / Publication No. US20030119018A1
 GENERAL INFORMATION:
 / APPLICANT: OMURA, SATOSHI
 / APPLICANT: IKEDA, HARUO
 / APPLICANT: ISHIKAWA, JUN
 / APPLICANT: HORIKAWA, HIROSHI
 / APPLICANT: SHIBA, TADAYOSHI
 / APPLICANT: SAKAKI, YOSHIYUKI
 / APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 / CURRENT APPLICATION NUMBER: US/10/156,761
 / CURRENT FILING DATE: 2002-05-29
 / PRIOR APPLICATION NUMBER: JP 2001-204089
 / PRIOR FILING DATE: 2001-05-30
 / PRIOR APPLICATION NUMBER: JP 2001-272657
 / PRIOR FILING DATE: 2001-08-02
 / NUMBER OF SEQ ID NOS: 15109
 / SEQ ID NO 3629
 / LENGTH: 11058
 / TYPE: DNA
 / ORGANISM: Streptomyces avermitilis
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: (1)..(11058)
 US-10-156-761-3629
 Query Match 5.8%; Score 246.2; DB 14; Length 11058;
 Best Local Similarity 51.5%; Pred. No. 3.7e-54;
 Matches 709; Conservative 0; Mismatches 638; Indels 30; Gaps 5;
 QY 1574 TGCAGGCGCTGTTCCGCGCGCGGCTGAGCAGCTGCCATGCACTCGCCGTTGTGTCG 1633
 DB 1469 TGCAGAGGTGTTCCAGAGAGAGTGGCGCGGAGACCCCGAGGGGCTCCGGTCTCTCG 1528
 QY 1634 CGCGAAGACGCTCAGTACGAAAGCTTTGCGCGCTTCCGCGGACTTGGCGCGCGC 1693
 DB 1529 AGGGAGCGATGTACGTAACGCGGAGCTGAACGCCCGGCCAACCGGATTCGCCACGCCC 1588
 QY 1694 TCCGAGAGAGGCGCAAGCCCGAACACATTGTGTGCGGTGTATGAGAAAGGCTGGG 1753
 DB 1589 TGATGCGAAGGGGTTGCGCGCGCGAGAGCTGTGCGGCTAGAGCTGGAAGCGGCGATCG 1648
 QY 1754 AGCAGATTGTCGCGGTTCTCGCGGTGCTGAGTCAAGCGCGGCTTACGTCGATGATG 1813
 DB 1649 AACTGATCCCGGCTTGTCTGCGGATCTCAAGTCCGGGTGCGCGATGTGCGCGTGAAC 1708
 QY 1814 CCGACTTACCGGCGGAGCGATATCACTTCTGATCATGTGATAGTAAAGCTGTGC 1873
 DB 1709 CGGCAACCGGCGGAGCGGATCGGTAATGTGTGAGGAGCCGACAGGCGCGGTGTGG 1768
 QY 1874 TGAAGAGCAAGTGTGATGAGCAACTGTATGAGCGCGCGGAGATTCAGAGCGGCTGTG 1933
 DB 1769 TGAC-----CACGCGGAGACGCGGATTTGTCAGCTGGA-gctgctgctcttgc 1819
 QY 1934 TGAGGAGGCGCGGCTGAGAGCGGAGCGGAGCAAGCTTCGATGATGATTCATGACAC 1993

Db 1820 AACGAGACGCGCGGACCTGCGCGCGCGGACCTCCGACCCGCGGCTGCCCGGACGCC 1879
Qy 1994 CTTGCGATCTGCGGATGTCATCTACCTCGGAGTCCAGAGGTTGCCAAGGGGATGA 2053
Db 1880 CCGAAGACCTGATCTACAGATCTACCTCGGCTCGACGGGCGCGGCGAAGGGGCTCG 1939
Qy 2054 TGATGATCTATCGGGGTCCTGCAACACATCTCGGACATCAACGAGCGCTTCGAATAG 2113
Db 1940 CCTTCACCCACACCAACGCTCGTCCGCTCTGAGACGGGCGAAGAGCATACGCTTCA 1999
Qy 2114 GCGCGGAGACAGAGGTCGCGGCTCTGCTCGCTGAGCTTCGATCTCTCGGCTATGATG 2173
Db 2000 CCGACACGAGCTCGGCGCGCTCTTCACAGCTACGCTTCGACGCTTCGCTCGGAGGA 2059
Qy 2174 TGTTCGAGATCTGCGCGCGGCGGCTGACGATCGTGTGCGCGGACGCTTCGAAGTGC 2233
Db 2060 TGTGGGGTGTCTGTCGACGGGCGCGCTGATGTCGTCCTTACGACGTCACCTCT 2119
Qy 2234 ATCCGCGCATTTGGGCAAGTTGATCGAACGAGAGAGTGAAGGTTGGAATTCGCTGC 2293
Db 2120 CCGCGAGAGATTCCTCGACCTGCTGTCGCGGACGCGTCACTGCTCAACGAGAGC 2179
Qy 2294 GCGGCTGATGCGATGCTGCTGAGCATTTTGAAGGTCGCGCGATTCGCTCGTACGT 2353
Db 2180 CTTCCGCTTCCTGCTGCTGCTGCTGCGCGGACGCGGACGCTGCTGCGCGGCTGCT 2239
Qy 2354 CTTGCGGCTTCGCTGCTGAGCGCGGACGCTGATCCGCGTGGGCTGCTGCGGAGCTGC 2413
Db 2240 TCGCGGGGAGAGTGGAGATCTCCGAACTGCGCGCTCGGCGGACGCGGTTCCGGCTCG 2299
Qy 2414 AGGCGATGAGCGCGCGCTGTCGATGATCAAGCTGGGCGGCGCAACGAGGTCGATCT 2473
Db 2300 ACCGATGCGCTG-----TCAATGTAAGGATCAACGAGACACCGCTC 2347
Qy 2474 GGTGATGAGGATCCCGCTGAGAACGTCGACT---ATCGTGGCGAGCATCCCTTACG 2530
Db 2348 ACACCACTTACCAACGCGCTGACCGAGCGGACGCTGAGCCCGGCGGCGCAACGCGGCTG 2407
Qy 2531 GCGGCTCGCTGCGCAACGAGAGTTCACAGTCTGATGAGGCGCTGCAACCGCGCGCG 2590
Db 2408 GCGATCCGCTGCGGACCTCGGGGCTGTAATCTCTGGAAGCGGACGAGACGTCGCTCG 2467
Qy 2591 TCTGGGTTCCGGGCAACTTACATTTGCGGCGGCTGCGCTGCGCTGAGCTGAGCTGAG 2650
Db 2468 TCGGGGTCGCGGTGATTCACAGTCCGCGCGCGCGGCTGCGCAACCTCAAC 2527
Qy 2651 ATGAAGAGAAACGCGCAAGCTCTCTGTCGACCCGCA-----GACCGGAGAGCGCC 2704
Db 2528 GCGCGGACTTACGCGCGAGCGGCTTCTCCGGAACGCTACGCTCGCGGCTGCTGCTC 2587
Qy 2705 TCTCAAGACGCGGATCTGCGGCGCTACCTGCGCGATGAAACATCGAGTTCAATGGGCG 2764
Db 2588 TCTACAGAGCGGCACTTCGCGCGCGCGCTGCGCGGAGCGGAGCTTGAAGTTCTGCGGCC 2647
Qy 2765 GTGAGGCAACCAATCAAGCTTCGCGGATACCGGCTGAGCTCGGCGGAAATCGAGAAA 2824
Db 2648 GAGTGAAGACAGGTCAGAAATCGCGGCTTCGAGTGAAGTGGCGAGATCGAGACG 2707
Qy 2825 CGCTCAAGTCCGATTCGAAGCTGACGAGCGGCTGATGTCGCTCGGAGACGAGCGG 2884
Db 2708 CGCTGGCGCGCATTCGAGGCTCGTGAAGCGCTGCTGCTGCTGCGAGGACACCGCG 2767
Qy 2885 CGAACAGCTCTCTTCAAGCTATGTCGCGGAGGCGACGAGAGCGCGCTGCG 2941
Db 2768 GCGACAGCGCTCTGCTGCTTACAGACGCGCGCGCGACGAGCCCGCGCGCTCG 2824

RESULT 13

US-09-940-316B-1
; Sequence 1, Application US/09940316B
; Publication No. US20030175501A1
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.

; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTU, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDES ENCODING THE *kdsA* GENE OF THE *PK-520* POLYKETIDE SYNTH
; TITLE OF INVENTION: GENE CLUSTER
; FILE REFERENCE: 30062-20026.11
; CURRENT APPLICATION NUMBER: US/09/940.316B
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 09/410,551
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275) ... (71465)
US-09-940-316B-1

Query Match 5.7%; Score 241.6; DB 12; Length 77536;
Best Local Similarity 49.7%; Pred. No. 8.5e-53;
Matches 703; Conservative 0; Mismatches 699; Indels 12; Gaps 3;

Qy 1530 GCGAGCGCAAGCGCGCAACGCGCTGCTGAGGAGATAGCTGCAACGCGGCTGTCG 1589
Db 49194 GGTGCTTCGAGACACAGCGCGGAGACTGCGCTGACAGCTGCGGCTGCTGCG 49253
Qy 1590 GCGCGGTCGAGGAGCTGCCATGACGCTGCGGCTGTCGCGCGCAAGAGCTTAC 1649
Db 49254 CCGGTACCGCGACGACACCCCGCGCGCGCTGACGACCGGACATCTCCCTAC 49313
Qy 1650 GTACGAAGCTTTCGCGCGCTTTCGCGGCACTTGGCGCGGCTGCGCGAGAGGGG 1709
Db 49314 CTAGCGCAGCTGAGACGCGCGCGGCAACCGCTGCGCACCTGCTCGCGCGCGGAC 49373
Qy 1710 ACGCCGCAACATTTGTCGCGGTGCTGATGAGAAAGCTGCGGAGCAGGTTGTCGCG 1769
Db 49374 GCGACCGCGGACCTGTCGCGGATCTGCGCGATGCGCGCGGACCTGATCTGCGAT 49433
Qy 1770 TCTGCGGTGCTGAGTCAGGCGGCGCTACGTCGATCGATCGGACTTACCGCGGA 1829
Db 49434 CGTGGGATCTTCAAGCGGCGCGCGCTTATGTCCTCGTGAACCCGAAACATCTCCGA 49493
Qy 1830 GCGTATCACTACTCTCTGATCATGATGAGTAAAGCTGTCGTCGACGACCATGCT 1889
Db 49494 GCGACGCGGCTTCTGTCGCGGACGCGACGCTACCAAGTGTGCGCA-----CGA 49547
Qy 1890 GAGTGAAGACTGTCATGCGCGCGGAGATCCAGCGCTGCTCTGAGCGGAGCGCGG 1949
Db 49548 GGTCTACCGTTCCGCGTTCGCGATGTCGCGACGCTGCTGCTGAGCGACCGGAGCT 49607
Qy 1950 CGAAGCGGACGCGGACGAGCTTCGATGATGCGCATTCAGACATTCGAGTCTCGCTGA 2009
Db 49608 GGAACGCGGACGCGGACGAGCGGCGCGGAGCTGAGTGAACCGGACACCTCGCTTA 49667
Qy 2010 TGTCTATCACTCTCGGATTCACAGGCTTTCGCAAGGCGGTGATGATCATTCGCGG 2069
Db 49668 GCGGATCTACAGCTTCGCGGTGACCGGAGGCGGAGCGGCTGCTACGCGGCTGTCAG 49727
Qy 2070 TGCGCTCAACATCTTGAACATCAACAGCGCTTGAATATAGGCTCCGAGACAGGCT 2129
Db 49728 GCGCGTCAACCTGCTGCTTGGAGAGCGGACGATGCGCGGCGGACCGGACCGGAC 49787


```

; LOCATION: (75535)..(76464)
; OTHER INFORMATION: ORF 23; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (78110)..(76449)
; OTHER INFORMATION: ORF 24; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (79864)..(78107)
; OTHER INFORMATION: ORF 25; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (81624)..(79861)
; OTHER INFORMATION: ORF 26; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (81909)..(81682)
; OTHER INFORMATION: ORF 27; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (82316)..(82062)
; OTHER INFORMATION: ORF 28; negative strandedness
; NAME/KEY: misc_feature
; LOCATION: (82587)..(8446)
; OTHER INFORMATION: ORF 29; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (84481)..(85548)
; OTHER INFORMATION: ORF 30; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (85556)..(86845)
; OTHER INFORMATION: ORF 31; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (87372)..(86803)
; OTHER INFORMATION: ORF 32; positive strandedness
; NAME/KEY: misc_feature
; LOCATION: (87494)..(88420)
; OTHER INFORMATION: ORF 33; positive strandedness, N-terminus only
US-09-976-059-1

```

```

Query Match      5.6%; Score 239; DB 10; Length 88421;
Beet Local Similarity 50.4%; Pred. No. 4.2e-52;
Matches 729; Conservative 0; Mismatches 695; Indels 23; Gaps 5;

```

```

QY 1525 GCGCGGCGGAGCGCAACGCGGCAACGCGCTGTGAGGAGCATACGCTGACGGCTTG 1584
DB 68266 GCAACCGGACCGGCAACGCTGCGCGCGCGCGCGCGCGCTGCGACGCTGACGGCTTG 68335
QY 1585 TTCCGCGGCGGCGGTGAGAGAGCTGCGCATGACGCTGCGCGGTGTGCGCGCGCAAGC 1644
DB 68326 TTCCGCGGAGAGGTGCGCGCTGTGCTCCACGCGCGGTGCGCTGCGAGATGGGGTGG 68385
QY 1645 CTCACGTACGAGAGAGCTTTCCGCGCGGTGCGCGCACTTGGCGCGCGCTGCGCGACG 1704
DB 68386 CACACCTTACCGGAGAGCTGACGAGGTGTCCGCGCGCTGCGCGCGGTGCGGAAAGGCC 68445
QY 1705 GGGGACCGCGGCAACATTTGTTGCGCGGTGTGATGAGAAAGGCTGGGAGAGATTGTC 1764
DB 68446 GCGGTCGCGAGCGGAGCGCGGTGCGCGCTGCTGACCGCTGCGCGAGCTGCTGCGC 68505
QY 1765 GCGGTTCTCGCGGTCTGAGTCAAGCGCGGCTGACGTGCGAGTATGATGATGATGATG 1824
DB 68506 GTGCTCTCTGCGGTCTCAAGCGCGCGCGCGCTGATGTGCGCTGACGCGCGGTGATCCG 68555
QY 1825 GCGGAGCGTATCACTACTCTCTGATCATGATGATGATGATGATGATGATGATGATG 1884
DB 68566 GCGGAGCGGATGCTTACACCTCTGCGCGAGCGCGCGCTGCGGTGTGATGATGATG 68625
QY 1885 TGGCGATGATGCAACTGTGATGCGCGCGCGGATGCAAGCGGCTGCTGTGAGAGAGCC 1944
DB 68626 GCGGATTTTCCGAGGTGAGCGGTGTGCGGTGCTGCGCGAGAGCTGCGCGAGAGCC 68685
QY 1945 GCGGTCGAGAGCGAGCGGAGCGGCTTCCGATGATGATGATGATGATGATGATGATG 2004
DB 68686 GCGGAGAGAGCGGAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATG 68736
QY 2005 GCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2064
DB 68737 GCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 68796

```

```

QY 2065 CCGGTCGCGTCAACACCATCTGAGATCAACGAGCGCTTGAATAGAGGCGCGAGAC 2124
DB 68797 GCGAAGGTGTGCGGCTGCTGAGAGCGGAGAGAGTATGCGCTGCGCGCGGAGC 68856
QY 2125 AGGTCGTGCGGCTCTCTGCTGAGCTTGCATCTCTGCTGATGATGATGATGATGATG 2184
DB 68857 GTGTGAGACCTTCTTCACTGCGCGCGCTTGCATCTCTGCTGAGATGATGAGGCTGC 68916
QY 2185 CTGCGCGGCGGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2244
DB 68917 CTGCTACCGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 68976
QY 2245 TGGGAGAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2304
DB 68977 TTTCAGAGACCTGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69036
QY 2305 CGGATGCTGCTGCGAGATTTTGAAGGTGCGCGCGGATGCTGCTGATGATGATGATG 2363
DB 69037 ACGGAGCTGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69096
QY 2364 TTCCGTCGTGAGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2423
DB 69097 TTCCGTCGTGAGCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69156
QY 2424 GCGCGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2483
DB 69157 GCGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69215
QY 2484 GTACCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2543
DB 69216 GATGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69275
QY 2544 CAACAGAGCTTCAACAGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 2603
DB 69276 GCGGTCGCGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 69335
QY 2604 GCAACTTCAACTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2663
DB 69336 GCAATCTACGTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69395
QY 2664 GCGGAGAGCTTCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2723
DB 69396 GCGGAGAGCTTCTGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69452
QY 2724 GCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2783
DB 69453 GCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69512
QY 2784 GCTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2843
DB 69513 GCTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69572
QY 2844 GCTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2894
DB 69573 GCTTGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69632
QY 2895 CTTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2954
DB 69633 GCTTACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69692
QY 2955 CTTCAAG 2961
DB 69693 CCGGATG 69699

```

```

RESULT 15
US-10-156-761-3630
; Sequence 3630, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO

```



```

APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 3630
LENGTH: 5127
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(5127)
US-10-156-761-3630

```

```

Query Match      5.5%; Score 233.4; DB 14; Length 5127;
Best Local Similarity 51.0%; Pred. No. 7.1e-51;
Matches 692; Conservative 0; Mismatches 641; Indels 24; Gaps 5;

```

```

QY 1559 TGAGCAGCATACGCTGCAAGCGCTGTCGCGCGCGGGTCGAGCAGTCCCATCAGC 1618
DB 2 TGAGCAGCGAAGCTTTTGGGAAAGCTTCCGAGAGAGGGGTTCCGTACGCGCAGCGG 61
QY 1619 TGGCGGTGTCGCGCGCGCGAGAGCGCTCAGCTGAGAGAGCTTTGCGCGCTTCCGCG 1678
DB 62 CCGCGCTGTCAGCAGGAGCGCGGACGCTGACGTAAGAGAGCTGAGCGTCCCTGCG 121
QY 1679 GACTTGGCGCGCGCTGCGCGAGCAGGAGGAGCAGCGCGAACAATGTCGCGGTGTA 1738
DB 122 GCATGCGCGCGCAGCTGCGGAGCTTGGGTGCGGAGCGCGAGAGCTGTCGCGGTAGCG 181
QY 1739 TGGAGAAAGCTGGAGAGAGCTTGTGCGCGCTTCTGCGGTGTCGAGTCAAGCGCGCT 1798
DB 182 TGGCGCGCGCGCTGAGTGTGATGTCGCGCTGCGCTGCGCGCGCGCGCGCGCGGCT 241
QY 1799 AGTGGCGGATTCATGCGCGAGCTACCGCGAGAGGTATCCACTGCTTCTGATGATGTG 1858
DB 242 ATGTCCGCTGAGACCGCGCGCAAGCGCGCGCGCTGTCGAGCTGCGCGCGAGGCGG 301
QY 1859 AGGTAAAGCTGCTGAGCGCAGCAGCATGCTGATGAGCAAACTGTATGCGCGCGGGA 1918
DB 302 GGGCGCGCTTGTGTGTCGCGAGCAGCGCGCTTGGCGGCG--CGGTGCGAGCGCGGTG 359
QY 1919 TCCAGCGGCTGCTGAGAGAGCGCGCGTTCGAGAGCGAGCGGCGACAGCGCTCCGATGA 1978
DB 360 ----CGGCGCGCTGAGACCGAGAGATCCGAGCGGTCTGAGACCGCGCGGCGCG 415
QY 1979 TGGCCATTAGACACTTTCGAGATCTGCGGTATGTCACTTACACTTGGGATCCAGGGT 2038
DB 416 TCCCGCGCGCGAGACCGCGCGCAAGCGCGCTACCGCGCTTTCACCTCCGTTTCAGCGGCG 475
QY 2039 TGGCCAAAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2098
DB 476 GGGCCAAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 535
QY 2099 AGCGCTTCAAAATAGGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2158
DB 536 AGCGGATGAGCGCTGCGGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 595
QY 2159 TCTCGGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2218
DB 596 CGGCGCGCTGAGAGATCTTTCGCGCGCGCTGTCGCGCGCGAGAGAGAGAGAGAGAGAG 655
QY 2219 CGTCAAGCTGCGGATTCGCGCGCGCATTTGGCAGAGTTGATCGAAGAGAGAGAGAGAG 2278

```

```

DB 656 CGGCGCGAGAGCGAGACCGCGCGCGCTGTCGCGCGGTGAGCGAGACAGCGGCTCACCG 715
QY 2279 TGTGAACCTCGGTGCGCGCGCGCTGATGCGGATGCTGTGAGCATTTTGAAGGTCCGCCG 2338
DB 716 TGTCTCAGGTGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 774
QY 2339 ATTGCTGCTGAGTCTGTCGCGCGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2398
DB 775 ----TGGAGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 829
QY 2399 TGTCTGCGAGCTTCCAGGCGCATCAGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 2458
DB 830 TGGCGCGGTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 889
QY 2459 CCGAAGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2518
DB 890 CCGAGTGTCTTCATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATC 946
QY 2519 GCATCCCTAGCGCGCGCTGCGCGCGCAAGAGCTTCCAGTGTGCTGCTGCTGCTGCTGCTG 2578
DB 947 CGGTCCCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1006
QY 2579 AACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2638
DB 1007 TCCCGGTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1066
QY 2639 GCTACTGCGCGCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2689
DB 1067 GCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1126
QY 2690 AGACCGGAGAGCGCTTCAAGAGCGCGCGATCTGGCGCGCTACCTGCGCGAGAGAGAG 2749
DB 1127 AGCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
QY 2750 TCGAGTTTCAAGGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2809
DB 1187 TCGAGTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1246
QY 2810 GGGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2869
DB 1247 CCGAGGTGAGAGCGCGCGCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1306
QY 2870 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2906
DB 1307 AGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1343

```

Search completed: October 4, 2003, 12:40:37
Job time : 1029.11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 3, 2003, 11:46:26 ; Search time 224.251 Seconds
(without alignment)
8331.640 Million cell updates/sec

Title: US-10-014-717-1_COPY_11872_16104

Perfect score: 4233
Sequence: 1 atgacgatacattcagctctc.....aggcgagagagagctcaaa 4233

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA.*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCIOS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4233	100.0	68750	US-09-335-409-1	Sequence 1, Appli
2	4233	100.0	68750	US-09-568-102-1	Sequence 1, Appli
3	4233	100.0	68750	US-09-567-969-1	Sequence 1, Appli
4	4233	100.0	68750	US-09-568-480-1	Sequence 1, Appli
5	4233	100.0	68750	US-09-568-486-1	Sequence 1, Appli
6	4233	100.0	68750	US-09-568-472-1	Sequence 1, Appli
7	4233	100.0	68750	US-09-567-899-1	Sequence 1, Appli
8	4114.6	97.2	71989	US-09-443-501A-2	Sequence 2, Appli
9	520.6	12.3	4563	US-09-252-991A-4765	Sequence 4765, Ap
10	491	11.6	2865	US-09-252-991A-6775	Sequence 4675, Ap
11	455	10.7	2922	US-09-252-991A-4815	Sequence 4815, Ap
12	358.2	8.5	4411529	US-09-103-840A-1	Sequence 1, Appli
13	349.2	8.2	4403765	US-09-103-840A-2	Sequence 2, Appli
14	348.6	8.2	18442	US-09-252-991A-6768	Sequence 4678, Ap
15	340	8.0	2844	US-09-252-991A-4764	Sequence 4764, Ap
16	274.4	6.5	3315	US-09-252-991A-8892	Sequence 8892, Ap
17	274.4	6.5	7911	US-09-252-991A-1182	Sequence 8892, Ap
18	261.4	6.2	4236	US-09-252-991A-7057	Sequence 7057, Ap
19	261.4	6.2	10023	US-09-252-991A-6997	Sequence 6997, Ap
20	246.6	5.8	6858	US-09-252-991A-1219	Sequence 1219, Ap
21	241.6	5.7	77536	US-09-410-551B-1	Sequence 1, Appli
22	240.8	5.7	1320	US-09-252-991A-4790	Sequence 4790, Ap
23	229.8	5.4	6573	US-09-252-991A-9183	Sequence 8888, Ap
24	229.6	5.4	1857	US-09-252-991A-8888	Sequence 8888, Ap
25	226.6	5.4	630	US-09-252-991A-8891	Sequence 8891, Ap
26	223.2	5.3	2025	US-09-252-991A-8891	Sequence 8891, Ap
27	220.2	5.2	7374	US-09-252-991A-9100	Sequence 9100, Ap

c	28	218.8	5.2	1689	4	US-09-252-991A-8717	Sequence 8717, Ap
c	29	202	4.8	4242	4	US-09-252-991A-7056	Sequence 7056, Ap
c	30	200.6	4.7	1965	4	US-09-252-991A-9230	Sequence 9230, Ap
c	31	194	4.6	2274	4	US-09-252-991A-6998	Sequence 6998, Ap
c	32	194	4.6	2823	4	US-09-252-991A-7100	Sequence 7100, Ap
c	33	194	4.6	3456	4	US-09-252-991A-7025	Sequence 7025, Ap
c	34	186	4.4	1509	4	US-09-252-991A-63	Sequence 63, Appl
c	35	186	4.4	7527	4	US-09-252-991A-71	Sequence 71, Appl
c	36	182.8	4.3	804	4	US-09-252-991A-4819	Sequence 4819, Ap
c	37	175.4	4.1	954	4	US-09-252-991A-4820	Sequence 4820, Ap
c	38	169.2	4.0	2805	4	US-09-252-991A-9064	Sequence 9064, Ap
c	39	168.2	4.0	2219	3	US-08-510-646B-17	Sequence 17, Appl
c	40	166.2	3.9	1494	4	US-09-252-991A-7049	Sequence 7049, Ap
c	41	163.2	3.9	765	4	US-09-252-991A-4721	Sequence 4721, Ap
c	42	158.2	3.7	1178	3	US-08-861-774E-91	Sequence 91, Appl
c	43	155	3.7	1200	3	US-08-861-774E-85	Sequence 85, Appl
c	44	152.8	3.6	1299	4	US-09-252-991A-7051	Sequence 7051, Ap
c	45	151.4	3.6	3126	4	US-09-252-991A-9066	Sequence 9066, Ap

ALIGNMENTS

RESULT 1									
US-09-335-409-1									
Sequence 1, Application US/09335409									
Patent No. 6121029									
GENERAL INFORMATION:									
APPLICANT: Schnupp, Thomas									
APPLICANT: Lignon, James									
APPLICANT: Molnar, Istvan									
APPLICANT: Zirkle, Ross									
APPLICANT: Cyr, Devon									
APPLICANT: Goelach, Joern									
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES									
FILE REFERENCE: 4-30582A									
CURRENT APPLICATION NUMBER: US/09/335,409									
CURRENT FILING DATE: 1999-06-17									
NUMBER OF SEQ ID NOS: 30									
SOFTWARE: Patentin Ver. 2.0									
SEQ ID NO 1									
LENGTH: 68750									
TYPE: DNA									
ORGANISM: Sorangium cellulosum									
US-09-335-409-1									
Query Match									
Best Local Similarity 100.0%; Score 4233; DB 3; Length 68750;									
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	ATGACGATCATCGCTTGAAGAGCTGAGCAGGAGGTATCAAGTGGCGCGAT	60						
Db	11872	ATGACGATCATCGCTTGAAGAGCTGAGCAGGAGGTATCAAGTGGCGCGAT	11931						
Qy	61	GGGAGGCGCTTCAGATACAGGCGCCCAAGAACCCCTGAACTGCTGCTGA	120						
Db	11932	GGGAGGCGCTTCAGATACAGGCGCCCAAGAACCCCTGAACTGCTGCTGA	11991						
Qy	121	ATTCGAGCACAAGACAGATCTTACGATGCTCTGTCAGAGACTCCCGCAATTC	180						
Db	11992	ATTCGAGCACAAGACAGATCTTACGATGCTCTGTCAGAGACTCCCGCAATTC	12051						
Qy	181	ATTCGAGCACAAGACAGATCTTACGATGCTCTGTCAGAGACTCCCGCAATTC	240						
Db	12052	ATTCGAGCACAAGACAGATCTTACGATGCTCTGTCAGAGACTCCCGCAATTC	12111						
Qy	241	TACTGCTGAGCGAGCAGGAGCTTACGATGCTCTGTCAGAGACTCCCGCAATTC	300						
Db	12112	TACTGCTGAGCGAGCAGGAGCTTACGATGCTCTGTCAGAGACTCCCGCAATTC	12171						
Qy	301	TACGACTGTACGATCTTCGAGCTGCGAGGCTGAGCGCGCTTTCGGAAGTCTGCGG	360						
Db	12172	TACGACTGTACGATCTTCGAGCTGCGAGGCTGAGCGCGCTTTCGGAAGTCTGCGG	12231						

QY 361 CCGCAGACATGCTTGGGGCCACACGCTCCCGACATGATGCAAGTATCCAGCTTAA 420
Db 12232 CCGCAGACATGCTTGGGGCCACACGCTCCCGACATGATGCAAGTATCCAGCTTAA 12291
QY 421 GTCGACCGCAGACATCGAGATCATCATCTGCGCGGGCTCGACCGGAGCACAACGGAAGG 480
Db 12292 GTCGACCGCAGACATCGAGATCATCATCTGCGCGGGCTCGACCGGAGCACAACGGAAGG 12351
QY 481 AGGCTGCTGCTTGGGAGATGGATGTGCAACCGCATCTATGACACGAGCGCCCTCG 540
Db 12352 AGGCTGCTGCTTGGGAGATGGATGTGCAACCGCATCTATGACACGAGCGCCCTCG 12411
QY 541 CTCCTACAGTGTGCGCGCTTGGGCTGAGAGAGCGGCAAAACCGTCTGCTGCTAGTATC 600
Db 12412 CTCCTACAGTGTGCGCGCTTGGGCTGAGAGAGCGGCAAAACCGTCTGCTGCTAGTATC 12471
QY 601 GATCTCATTAAGTTAGCTAGGACGCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 660
Db 12472 GATCTCATTAAGTTAGCTAGGACGCTGTCCATCATCTTCAAGGACTGGCTCAGCTTC 12531
QY 661 TACGAAGATCCGAGACCTCTCTCCCTGTCTGAGCTCTGTAACCGCATTAATGATCTC 720
Db 12532 TACGAAGATCCGAGACCTCTCTCCCTGTCTGAGCTCTGTAACCGCATTAATGATCTC 12591
QY 721 GGGCTGAGATCTCGCAAGAGTCTGAGGCGCATCAACGATCGATGATTAATGGAAGCG 780
Db 12592 GGGCTGAGATCTCGCAAGAGTCTGAGGCGCATCAACGATCGATGATTAATGGAAGCG 12651
QY 781 CGCATCGCGAGCTCCACCTCCGCGGACGCTTCGATGAAGGCGCATCATCACTTC 840
Db 12652 CGCATCGCGAGCTCCACCTCCGCGGACGCTTCGATGAAGGCGCATCATCACTTC 12711
QY 841 AAGAGATCCGCTTCGCGACACAGAGCAATGGCTGCGGCTCTGCGGCTCGATTG 900
Db 12712 AAGAGATCCGCTTCGCGACACAGAGCAATGGCTGCGGCTCTGCGGCTCGATTG 12771
QY 901 AAGCGCGCTGCGGGAGCGCGGGCTGACCCCGACGCGGCTCATCTGCTGCTGATTTCC 960
Db 12772 AAGCGCGCTGCGGGAGCGCGGGCTGACCCCGACGCGGCTCATCTGCTGCTGATTTCC 12831
QY 961 GAGGTATCGGGCTGAGACGCGGACCCCGGTTTACGCTCAATAAGCTCTTCAAC 1020
Db 12832 GAGGTATCGGGCTGAGACGCGGACCCCGGTTTACGCTCAATAAGCTCTTCAAC 12891
QY 1021 CGGCTCCCGCTCATCGCGCGTGAACGATATCACCGGGGACTTCACTGATGCTCTC 1080
Db 12892 CGGCTCCCGCTCATCGCGCGTGAACGATATCACCGGGGACTTCACTGATGCTCTC 12951
QY 1081 CTGGAATGACACCACTCGCGACAGAGCTTGAACAGCGGCTTAAGGATTTCAAGAG 1140
Db 12952 CTGGAATGACACCACTCGCGACAGAGCTTGAACAGCGGCTTAAGGATTTCAAGAG 13011
QY 1141 CAGCTGTGGGAAGCGATGATCACTGCGAGTAAAGGATTCAGGCTCAGCGAGGCG 1200
Db 13012 CAGCTGTGGGAAGCGATGATCACTGCGAGTAAAGGATTCAGGCTCAGCGAGGCG 13071
QY 1201 GCCCGGGTCTGGGAGTCAACGAGCGCATTTGTTCCCGTGTGCTCAAGACGCGCTT 1260
Db 13072 GCCCGGGTCTGGGAGTCAACGAGCGCATTTGTTCCCGTGTGCTCAAGACGCGCTT 13131
QY 1261 AACCGCAAGTCTGTTGTCACCTGTTGACAGAGCTCGGAACCTCGGTTTACACAC 1320
Db 13132 AACCGCAAGTCTGTTGTCACCTGTTGACAGAGCTCGGAACCTCGGTTTACACAC 13191
QY 1321 ACGCAGACTCTGAGCTGCTGATCATCAGCTCTACGAGACGATGGGACCTGCTC 1380
Db 13192 ACGCAGACTCTGAGCTGCTGATCATCAGCTCTACGAGACGATGGGACCTGCTC 13251
QY 1381 CTCGCTGGGACATGCTGACGAGAGTGTTCGCGCCGACCTTCTGACGACATGCTGAA 1440
Db 13252 CTCGCTGGGACATGCTGACGAGAGTGTTCGCGCCGACCTTCTGACGACATGCTGAA 13311

QY 1441 GGTAGCTGTTTTCTCCGGGCGCTCACTAGAGAACCATGGGAGTGAACAGGTGCGCTGT 1500
Db 13312 GGTAGCTGTTTTCTCCGGGCGCTCACTAGAGAACCATGGGAGTGAACAGGTGCGCTGT 13371
QY 1501 TCGCTTCGCGCTGCCAGCTAGAAAGCGGGCGAGCGCAAAACGCAACCAACGCGCTGCTG 1560
Db 13372 TCGCTTCGCGCTGCCAGCTAGAAAGCGGGCGAGCGCAAAACGCAACCAACGCGCTGCTG 13431
QY 1561 ACGGACATAGCGCTGACAGCGCTGTTCCGCGCGCGGGTTCAGACAGCTGCCATGACCTC 1620
Db 13432 ACGGACATAGCGCTGACAGCGCTGTTCCGCGCGCGGGTTCAGACAGCTGCCATGACCTC 13491
QY 1621 GCCGCTGCTGGCGCGGAGACGCTCACTAGAGAGCTTTGCGCGCGCTTCCGCGGCGGA 1680
Db 13492 GCCGCTGCTGGCGCGGAGACGCTCACTAGAGAGCTTTGCGCGCGCTTCCGCGGCGGA 13551
QY 1681 CTTGGCGCGCGCTGCGGAGAGGAGGACGCGCCGGAACATTTGCTGCGGTGATGATG 1740
Db 13552 CTTGGCGCGCGCTGCGGAGAGGAGGACGCGCCGGAACATTTGCTGCGGTGATGATG 13611
QY 1741 GAGAAAGCTGGGAGACAGGTTGTCCGCGTCTGAGTCAAGGCGCGCTTAC 1800
Db 13612 GAGAAAGCTGGGAGACAGGTTGTCCGCGTCTGAGTCAAGGCGCGCTTAC 13671
QY 1801 GTGCGGATCGATGCGGACCTACCGGCGAGGATCTACCTCTCGATCATGATGAG 1860
Db 13672 GTGCGGATCGATGCGGACCTACCGGCGAGGATCTACCTCTCGATCATGATGAG 13731
QY 1861 GTAAAGCTGCTGCTAGCAGCAGCATGCTGATGCAAACTGTCAATGACGCGCGGGATC 1920
Db 13732 GTAAAGCTGCTGCTAGCAGCAGCATGCTGATGCAAACTGTCAATGACGCGCGGGATC 13791
QY 1921 CAGCGGCTGCTGCTAGGAGAGCGCGCGCTCGAAGCGCACGCGACCTTCCATGATG 1980
Db 13792 CAGCGGCTGCTGCTAGGAGAGCGCGCGCTCGAAGCGCACGCGACCTTCCATGATG 13851
QY 1981 CCCATTCAAGACCTTGGGATCTGCGGATGATCTCACTCGAGGATTCACAGGTTG 2040
Db 13852 CCCATTCAAGACCTTGGGATCTGCGGATGATCTCACTCGAGGATTCACAGGTTG 13911
QY 2041 CCCAAGGGGATGATGATGATCATGCGGGGTGCGGTCAACACCATCTTGAGACATCAACGAG 2100
Db 13912 CCCAAGGGGATGATGATGATCATGCGGGGTGCGGTCAACACCATCTTGAGACATCAACGAG 13971
QY 2101 CGCTTCGAATAGGCGCGCGAGACAGGGTCTGCGCTCTCTCGCTGAGCTTCAATCTC 2160
Db 13972 CGCTTCGAATAGGCGCGCGAGACAGGGTCTGCGCTCTCTCGCTGAGCTTCAATCTC 14031
QY 2161 TCGGCTATGATGTTGCGGATCTGCGCGCGGCGGTAAGATGCTGCTGCGGACGCG 2220
Db 14032 TCGGCTATGATGTTGCGGATCTGCGCGCGGCGGTAAGATGCTGCTGCGGACGCG 14091
QY 2221 TCCAAAGCTGCGGATCGCGCGATTTGGCAGAGTTGATCGAACGAGAGGTGACGCTG 2280
Db 14092 TCCAAAGCTGCGGATCGCGCGATTTGGCAGAGTTGATCGAACGAGAGGTGACGCTG 14151
QY 2281 TGGAACTGCTGCGCGCGCTGATGCTGCTGACAGCATTTTGAAGGTGCGCCCGAT 2340
Db 14152 TGGAACTGCTGCGCGCGCTGATGCTGCTGACAGCATTTTGAAGGTGCGCCCGAT 14211
QY 2341 TCGCTGCTAGGATCTGCGCGCTTTCGCTGACAGCGGACTGATCCCGTGGGCTG 2400
Db 14212 TCGCTGCTAGGATCTGCGCGCTTTCGCTGACAGCGGACTGATCCCGTGGGCTG 14271
QY 2401 CTTGGCGAGCTTCAGAGCATCAAGGCGCGGCTGTGATCATAGCTTGGCGGCGGAC 2460
Db 14272 CTTGGCGAGCTTCAGAGCATCAAGGCGCGGCTGTGATCATAGCTTGGCGGCGGAC 14331
QY 2461 GAAAGCTGATTTGCTGCTAGGAGTACCCCGTGAAGAACTGTGACATCTGCTGCGGAGC 2520
Db 14332 GAAAGCTGATTTGCTGCTAGGAGTACCCCGTGAAGAACTGTGACATCTGCTGCGGAGC 14391
QY 2521 ATCCCTAGAGCGGCTGCGGCGGAGAACGAGCTTCAAGTCTGATGAGGCGCTGAA 2580

```

Db 14392 ATCCCTACGCGCGCTCCGCTGGCAACGACGTTCCAGTCTGATGAGCGCTCCGAA 14451
Oy 2581 CGCGCGCGGCTCTGGGTTCCGGGGGCACTTATTTGGGGGGGGTGGGCTTGGCACTGGGC 2640
Db 14452 CGCGCGCGGCTCTGGGTTCCGGGGGCACTTATTTGGGGGGGGTGGGCTTGGCACTGGGC 14511
Oy 2641 TACTGGCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
Db 14512 TACTGGCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14571
Oy 2701 CGCCTTACAGAGCCGGGATTTGGGCGGCTTACTGCGGATGAAACATGAGTTCAATG 2760
Db 14572 CGCCTTACAGAGCCGGGATTTGGGCGGCTTACTGCGGATGAAACATGAGTTCAATG 14631
Oy 2761 GGGCGTGAAGCAACCAATCAAGCTTGGCGGATACCGGTTGAGCTGGGGGAATTCAG 2820
Db 14632 GGGCGTGAAGCAACCAATCAAGCTTGGCGGATACCGGTTGAGCTGGGGGAATTCAG 14691
Oy 2821 GAAACGCTCAAGTCGATCCGAACTAGCGGAGCGGCTGATTTGTCCTCGGAGAGAG 2880
Db 14692 GAAACGCTCAAGTCGATCCGAACTAGCGGAGCGGCTGATTTGTCCTCGGAGAGAG 14751
Oy 2881 GCGCGCAACAACTCTTCTTACCTTATGTGTCTCCGAGAGGCAACGAGAGCGCTGCC 2940
Db 14752 GCGCGCAACAACTCTTCTTACCTTATGTGTCTCCGAGAGGCAACGAGAGCGCTGCC 14811
Oy 2941 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
Db 14812 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14871
Oy 3001 GACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
Db 14872 GACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14931
Oy 3061 GACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120
Db 14932 GACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 14991
Oy 3121 GACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180
Db 14992 GACGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15051
Oy 3181 GAGTTTGTGATCTCTGAGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Db 15052 GAGTTTGTGATCTCTGAGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15111
Oy 3241 AAATTCGTTATCCATCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
Db 15112 AAATTCGTTATCCATCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15171
Oy 3301 TCCGCGCGCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3360
Db 15172 TCCGCGCGCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15231
Oy 3361 TTGCTGAAGAGCTCTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Db 15232 TTGCTGAAGAGCTCTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15291
Oy 3421 GTGTTGATGAAGCGGCTTCAACTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3480
Db 15292 GTGTTGATGAAGCGGCTTCAACTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 15351
Oy 3481 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3540
Db 15352 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 15411
Oy 3541 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3600
Db 15412 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 15471
Oy 3601 GAAAGGTTTGGCGCGGCTTCTGAGCTGAGCAATTCGAGCTTACGTCAGCGAGTGTG 3660

```

```

Db 15472 GAAAGGTTTGGCGCGGCTTCTGAGCTGAGCAATTCGAGCGTTTACGTCAGCGAGTGTG 15531
Oy 3661 GGGCGGCGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3720
Db 15532 GGGCGGCGGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15591
Oy 3721 AGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780
Db 15592 AGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15651
Oy 3781 CGGCACTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Db 15652 CGGCACTTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15711
Oy 3841 GATGAGTTTGGCGCGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 15712 GATGAGTTTGGCGCGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15771
Oy 3901 GATACCTGTCGCGCGGAGATTTGGGGGCAACGCGCGGAGAGAGAGAGAGAGAGAGATC 3960
Db 15772 GATACCTGTCGCGCGGAGATTTGGGGGCAACGCGCGGAGAGAGAGAGAGAGAGATC 15831
Oy 3961 CTGCTCGGCTGTCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4020
Db 15832 CTGCTCGGCTGTCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15891
Oy 4021 GTCGATCTTGGTGGAGATGATTCACATGCTTGGCATGAGAGAGAGAGAGAGAGAGAG 4080
Db 15892 GTCGATCTTGGTGGAGATGATTCACATGCTTGGCATGAGAGAGAGAGAGAGAGAGAG 15951
Oy 4081 CTGATAGAGAGAGATGCGCATACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Db 15952 CTGATAGAGAGAGATGCGCATACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16011
Oy 4141 TCCGTTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
Db 16012 TCCGTTTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16071
Oy 4201 GAGGTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4233
Db 16072 GAGGTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16104

```

```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patient No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1
Query Match 100.0%; Score 4233; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 ATGACGATCAATCAGCTTCTGAGAGAGCTCGAGCAGAGGATATCAAGCTTGGCGGCGGAT 60

```

Db 11872 ATGACATTAATAGCTTCTGAACAGCTCGAGCACAGGATATCAAGCTGGGCGCAT 11931
Qy 61 GGGAGCGCTCCAGATATAGGCGCCCAAGAAAGCCCTTGAACCCGAACCTGCTCGCA 120
Db 11932 GGGAGCGCTCCAGATATAGGCGCCCAAGAAAGCCCTTGAACCCGAACCTGCTCGCA 11991
Qy 121 ATCTCGAGACAAAGACAGATCTGACGATGCTCTGTCAGAGATCTCCCGCAGATCC 180
Db 11992 ATCTCGAGACAAAGACAGATCTGACGATGCTCTGTCAGAGATCTCCCGCAGATCC 12051
Qy 181 ATCGGCGCGCCCGCAGCGAGCGGCGCTCGCTTCTCTCAAGACATCCAGAAATCC 240
Db 12052 ATCGGCGCGCCCGCAGCGAGCGGCGCTCGCTTCTCTCAAGACATCCAGAAATCC 12111
Qy 241 TACTGCTGGGCGGACAGAGCGTTTACGATCCCGAGCGGATCCAGCCTATCGCA 300
Db 12112 TACTGCTGGGCGGACAGAGCGTTTACGATCCCGAGCGGATCCAGCCTATCGCA 12171
Qy 301 TACGACTGATCGGATTCGACGTGCGGAGCTGAGCGCGCTTTGGAAGTCTGCGG 360
Db 12172 TACGACTGATCGGATTCGACGTGCGGAGCTGAGCGCGCTTTGGAAGTCTGCGG 12231
Qy 361 CGGACAGACATGCTTGGGCGCCACAGCTGCCGACATGATGAGGATTCGAGCCTAA 420
Db 12232 CGGACAGACATGCTTGGGCGCCACAGCTGCCGACATGATGAGGATTCGAGCCTAA 12291
Qy 421 GTGACCGCGACATCGAGATCATCATGCTGCGGCGCTGACCGGAGCAACGCGAAGCG 480
Db 12292 GTGACCGCGACATCGAGATCATCATGCTGCGGCGCTGACCGGAGCAACGCGAAGCG 12351
Qy 481 AGGCTGCTGCTGTTGGAGATGCGATGTGCAACCGATCTATGACACCGAGCGCTCCG 540
Db 12352 AGGCTGCTGCTGTTGGAGATGCGATGTGCAACCGATCTATGACACCGAGCGCTCCG 12411
Qy 541 CTCTACACGCTGCTGCGCGCTGCGCTGACAGCGGCAACCGCTCTCGTCTCACTATC 600
Db 12412 CTCTACACGCTGCTGCGCGCTGCGCTGACAGCGGCAACCGCTCTCGTCTCACTATC 12471
Qy 601 GATCTCATTAAGCTTGAAGCTGAGCAGCTGTCCATCATCTTCAAGGACTGCGTCACTTC 660
Db 12472 GATCTCATTAAGCTTGAAGCTGAGCAGCTGTCCATCATCTTCAAGGACTGCGTCACTTC 12531
Qy 661 TACGAAGATCCGAGACCTTCTCTGCTGCTGAGACTCTGTAACCGGCTATATGATC 720
Db 12532 TACGAAGATCCGAGACCTTCTCTGCTGCTGAGACTCTGTAACCGGCTATATGATC 12591
Qy 721 GCGCTGAGCTGCGAAGAGCTGAGGCGCATCAAGATCGATGATTACTGGAAGCG 780
Db 12592 GCGCTGAGCTGCGAAGAGCTGAGGCGCATCAAGATCGATGATTACTGGAAGCG 12651
Qy 781 CGCATGCGAGAGCTCCACCTCCGCGAGCGCTTCCGATGAAGGCGCATCATCTACCTG 840
Db 12652 CGCATGCGAGAGCTCCACCTCCGCGAGCGCTTCCGATGAAGGCGCATCATCTACCTG 12711
Qy 841 AAGGAGATCCGCTTCCGCGACACAGGACATGCTGCGTCCGATCTCTGGGGTCCATTG 900
Db 12712 AAGGAGATCCGCTTCCGCGACACAGGACATGCTGCGTCCGATCTCTGGGGTCCATTG 12771
Qy 901 AAGCGGCTGTCGAGGAGCGGCGCTGACCCCGAGCGGCTATCTGCTGCTGATTTTC 960
Db 12772 AAGCGGCTGTCGAGGAGCGGCGCTGACCCCGAGCGGCTATCTGCTGCTGATTTTC 12831
Qy 961 GAGGTATCGGCGCTGAGCGGAGCGGCGCTTACGCTCAATCAACGCTCTTCAAC 1020
Db 12832 GAGGTATCGGCGCTGAGCGGAGCGGCGCTTACGCTCAATCAACGCTCTTCAAC 12891
Qy 1021 CGGCTCCCGCTCATCCGCGGTAACGATATACCGGGGACTTCAAGTGTGATCTC 1080
Db 12892 CGGCTCCCGCTCATCCGCGGTAACGATATACCGGGGACTTCAAGTGTGATCTC 12951
Qy 1081 CTGGAATGACACCACTCGCGACAAGAGCTTGAACAGCGGCTATTAAGAG 1140

Db 12952 CTGACATGACACCACTCGGAGACAGGCTTCCAAAGCGCGCTAAGCTATTCAAG 13011
Qy 1141 CAGCTGTGGAAAGCATGATATCATGCGACTGAAGCGGTATTCGAGTCCAGCGAGGCC 1200
Db 13012 CAGCTGTGGAAAGCATGATATCATGCGACTGAAGCGGTATTCGAGTCCAGCGAGGCC 13071
Qy 1201 GCGCGGGTCTGAGGATCCAAAGAGCGGCTATTTCCCGTGTGCTCAGGCGGCTT 1260
Db 13072 GCGCGGGTCTGAGGATCCAAAGAGCGGCTATTTCCCGTGTGCTCAGGCGGCTT 13131
Qy 1261 AACCAAGATGTTGTGTCACTTGTGAGAGGCTCGGAACCTCCGATGTACACAGC 1320
Db 13132 AACCAAGATGTTGTGTGTACCTGTTGAGAGGCTCGGAACCTCCGATGTACACAGC 13191
Qy 1321 ACGAGACTCTCAGCTGCTGTGATATCATGCTTACAGACAGATGGGAGCTTCTGTC 1380
Db 13192 ACGAGACTCTCAGCTGCTGTGATATCATGCTTACAGACAGATGGGAGCTTCTGTC 13251
Qy 1381 CTGCGGTGGACATGTCGAGGAGGTGTTCCCGCCGACCTTCTGGAAGACATGCTCGA 1440
Db 13252 CTGCGGTGGACATGTCGAGGAGGTGTTCCCGCCGACCTTCTGGAAGACATGCTCGA 13311
Qy 1441 GGTAGCTGCTTCTCTCGGCGCTCATGAGAACATGAGGAGTGAACAGGTGCTGT 1500
Db 13312 GGTAGCTGCTTCTCTCGGCGCTCATGAGAACATGAGGAGTGAACAGGTGCTGT 13371
Qy 1501 TCGCTTCCGCTGCTCCAGCTAGAGAGCGGCGGAGCGCAACGCAACGAGCTGCTG 1560
Db 13372 TCGCTTCCGCTGCTCCAGCTAGAGAGCGGCGGAGCGCAACGCAACGAGCTGCTG 13431
Qy 1561 AGCGAGATPAGCTGACAGGCGCTGTTCCGCGGCGGCTGAGACAGCTGCCATGAGCTC 1620
Db 13432 AGCGAGATPAGCTGACAGGCGCTGTTCCGCGGCGGCTGAGACAGCTGCCATGAGCTC 13491
Qy 1621 GCGGTGTGTGCGCGGAGAGCGCTCAAGTGAAGAGCTTTCGCGCTTCCGCGG 1680
Db 13492 GCGGTGTGTGCGCGGAGAGCGCTCAAGTGAAGAGCTTTCGCGCTTCCGCGG 13551
Qy 1681 CTGCGCGCGGCTGCGGAGAGCGGCGGAGCGCCGCAACATTTGTTGCGGTGTGATG 1740
Db 13552 CTGCGCGCGGCTGCGGAGAGCGGCGGAGCGCCGCAACATTTGTTGCGGTGTGATG 13611
Qy 1741 GAGAAAGCTGGAGCAGGTTGTGCGGTTCTCGGCTGCTGAGTCAAGCGCGGCTAC 1800
Db 13612 GAGAAAGCTGGAGCAGGTTGTGCGGTTCTCGGCTGCTGAGTCAAGCGCGGCTAC 13671
Qy 1801 GTGCCGATCGATGCGGACTTACCGGCGAGGCTATCACTACCTCTGATCATGTGAG 1860
Db 13672 GTGCCGATCGATGCGGACTTACCGGCGAGGCTATCACTACCTCTGATCATGTGAG 13731
Qy 1861 GTAAAGCTGTGCTGACGAGCATGCTGATGAGCAAACTGTCAATGCGCGCGGGATC 1920
Db 13732 GTAAAGCTGTGCTGACGAGCATGCTGATGAGCAAACTGTCAATGCGCGCGGGATC 13791
Qy 1921 CAGCGGCTGCTGTGAGCGAGGCGGCGCTGCAAGGCGAGCGGACCAAGCTTCCATGATG 1980
Db 13792 CAGCGGCTGCTGTGAGCGAGGCGGCGCTGCAAGGCGAGCGGACCAAGCTTCCATGATG 13851
Qy 1981 CCCATTAGACACTTGGATCTGCGATGTGATCTACCTCGGAGTCCACAGGGTTG 2040
Db 13852 CCCATTAGACACTTGGATCTGCGATGTGATCTACCTCGGAGTCCACAGGGTTG 13911
Qy 2041 CCCAAGGGGTGATGATGATATGAGGGGTGCGGTCAACCATCTGAGCATCAACAG 2100
Db 13912 CCCAAGGGGTGATGATGATATGAGGGGTGCGGTCAACCATCTGAGCATCAACAG 13971
Qy 2101 CGCTTGAAATGAGGCGCGAGAGAGGGTGTGCGGCTCTCTCGCTGAGCTTCCATCTC 2160
Db 13972 CGCTTGAAATGAGGCGCGAGAGAGGGTGTGCGGCTCTCTCGCTGAGCTTCCATCTC 14031
Qy 2161 TCGGCTATGATGTGTCGGAATCTTGCGGCGGCGGAGTCAATGCTGTGCGGAGCGG 2220
Db 14032 TCGGCTATGATGTGTCGGAATCTTGCGGCGGCGGAGTCAATGCTGTGCGGAGCGG 14091

QY 2221 TCCAGCTGCGGATCCGGCGATTGAGCAGAGTTGATCGAACGAGAGAGTGA CGGTG 2280
DB 14092 TCCAGCTGCGGATCCGGCGATTGAGCAGAGTTGATCGAACGAGAGAGTGA CGGTG 14151
QY 2281 TGGAACTGGGTGCGGCGCTGATGCGGATGCTCGTGCAGCATTTTGAAGGTGCGCCGAT 2340
DB 14152 TGGAACTGGGTGCGGCGCTGATGCGGATGCTCGTGCAGCATTTTGAAGGTGCGCCGAT 14211
QY 2341 TGGCGCTAGGATCTGGGCTTTGCTGCTGAGCGCGCATGGAATCCCGTGGGCGCTG 2400
DB 14212 TGGCGCTAGGATCTGGGCTTTGCTGCTGAGCGCGCATGGAATCCCGTGGGCGCTG 14271
QY 2401 CCTGGCGAGCTTCAGAGCCATAGAGCCCGCGCTGCTGATGATGAGCTGGGCGGCGCAC 2460
DB 14272 CCTGGCGAGCTTCAGAGCCATAGAGCCCGCGCTGCTGATGATGAGCTGGGCGGCGCAC 14331
QY 2461 GAAGCTGATCTGGTGCATCGGGTACCCCGTGAAGAACTGACCTATCTGGGCGAGC 2520
DB 14332 GAAGCTGATCTGGTGCATCGGGTACCCCGTGAAGAACTGACCTATCTGGGCGAGC 14391
QY 2521 ATCCCTAGCGCGCTGCGCTGCGCAACGAGCTTCCAGCTGCTGATGAGAGCGCTCGAA 2580
DB 14392 ATCCCTAGCGCGCTGCGCTGCGCAACGAGCTTCCAGCTGCTGATGAGAGCGCTCGAA 14451
QY 2581 CCGCGCCGCTGCGGCTTCCGGGCGAATCTACATTTGGCGGGTCTGGCACTGGGAC 2640
DB 14452 CCGCGCCGCTGCGGCTTCCGGGCGAATCTACATTTGGCGGGTCTGGCACTGGGAC 14511
QY 2641 TACTGGCGGATGAGAGAGAGACGCGCAAGCTTCTCTGTCACCCCGGAGACCGGGAG 2700
DB 14512 TACTGGCGGATGAGAGAGAGACGCGCAAGCTTCTCTGTCACCCCGGAGACCGGGAG 14571
QY 2701 CGCCTCTACAGACCGGCGATCTGGGCGGCTACCTGCGCGATGGAATGAGATTCATG 2760
DB 14572 CGCCTCTACAGACCGGCGATCTGGGCGGCTACCTGCGCGATGGAATGAGATTCATG 14631
QY 2761 GGGCGTGAGACAACCAATCAAGCTTGGCGGATACCGGCTGAGCTGGGGAATTCAG 2820
DB 14632 GGGCGTGAGACAACCAATCAAGCTTGGCGGATACCGGCTGAGCTGGGGAATTCAG 14691
QY 2821 GAAACGCTCAAGTGCATCCGAACGTACCGGACGCGGTGATTTGCGCTCGGAGAC 2880
DB 14692 GAAACGCTCAAGTGCATCCGAACGTACCGGACGCGGTGATTTGCGCTCGGAGAC 14751
QY 2881 GCGCGAACAAGCTCTTCTAGCTTATGCTTCCCGGAGGCGACACGAGACGCGCTCC 2940
DB 14752 GCGCGAACAAGCTCTTCTAGCTTATGCTTCCCGGAGGCGACACGAGACGCGCTCC 14811
QY 2941 GAGCAGACGCGAGCTCAAGACCGAGCGGATCGACGAGAGACACGCGCGGAGACG 3000
DB 14812 GAGCAGACGCGAGCTCAAGACCGAGCGGATCGACGAGAGACACGCGCGGAGACG 14871
QY 3001 GACGCTTGAAGCGAGCGGAGAGGATGACGTTCAAGCTCGCTCGACACGAGCTCCGAG 3060
DB 14872 GACGCTTGAAGCGAGCGGAGAGGATGACGTTCAAGCTCGCTCGACACGAGCTCCGAG 14931
QY 3061 GACCTGACGGAAGCCGCTGCTGATCTGACCCGCGGAGATCCGCGGAGCGCGGCTG 3120
DB 14932 GACCTGACGGAAGCCGCTGCTGATCTGACCCGCGGAGATCCGCGGAGCGCGGCTG 14991
QY 3121 GACGCTGACGCGCTGCGCTGATGAGCTCGAAAGTTCTTTGAGAGCCCGGATTCGTTGTT 3180
DB 14992 GACGCTGACGCGCTGCGCTGATGAGCTCGAAAGTTCTTTGAGAGCCCGGATTCGTTGTT 15051
QY 3181 GAGTTTGTGATTCCTGAGCTGCTTGAAGACGCTGAGACCGCGGAGCGCGGCTTCC 3240
DB 15052 GAGTTTGTGATTCCTGAGCTGCTTGAAGACGCTGAGACCGCGGAGCGCGGCTTCC 15111
QY 3241 AAATTCGTTATCCATCGCGGCGGAGACGTAACCGGCTGCAAACTAGCGGATTCGAA 3300
DB 15112 AAATTCGTTATCCATCGCGGCGGAGACGTAACCGGCTGCAAACTAGCGGATTCGAA 15171

QY 3301 TCCGCGCATCGAGGCGGTGAGACGAGGCTTCTATTATTAACAACCGCTTCGAGACCGT 3360
DB 15172 TCCGCGCATCGAGGCGGTGAGACGAGGCTTCTATTATTAACAACCGCTTCGAGACCGT 15231
QY 3361 TTGCTGAAGCTCTCCGATTCACGGGATCGAGGCGGAGCGGACGTTCCGCAAACTTCAC 3420
DB 15232 TTGCTGAAGCTCTCCGATTCACGGGATCGAGGCGGAGCGGACGTTCCGCAAACTTCAC 15291
QY 3421 GTGTTTCAAGAGCGGCTTCAACCTCTGTTCTGGGACAGGATCGAGCCATGAGTGC 3480
DB 15292 GTGTTTCAAGAGCGGCTTCAACCTCTGTTCTGGGACAGGATCGAGCCATGAGTGC 15351
QY 3481 CTGATGAGATCGTGTGCGGAAATTTTCCCTGCTGAGGCGGATATATGCGGAGCTC 3540
DB 15352 CTGATGAGATCGTGTGCGGAAATTTTCCCTGCTGAGGCGGATATATGCGGAGCTC 15411
QY 3541 CTGATGAGACGCGGCTTCTGCAACATCGGCGTCTGTCCGCTGGGCAATTCATTTT 3600
DB 15412 CTGATGAGACGCGGCTTCTGCAACATCGGCGTCTGTCCGCTGGGCAATTCATTTT 15471
QY 3601 GAACAGGTTGGCGGCTTCTGACCTGCACTTCGAGCGTTTACGTCACGCGCATGCTG 3660
DB 15472 GAACAGGTTGGCGGCTTCTGACCTGCACTTCGAGCGTTTACGTCACGCGCATGCTG 15531
QY 3661 GCGCGCGGCTGAGACCGCGGAGTTCAGGCTGTGACGCTCGGTCAAGATTCCTCACCG 3720
DB 15532 GCGCGCGGCTGAGACCGCGGAGTTCAGGCTGTGACGCTCGGTCAAGATTCCTCACCG 15591
QY 3721 AGCGCGGCTGAGACCGCGGCGGCTTCCGCGCGGAGACGACCTTGGCGATTCGTT 3780
DB 15592 AGCGCGGCTGAGACCGCGGCGGCTTCCGCGCGGAGACGACCTTGGCGATTCGTT 15651
QY 3781 CGGCACTTCTTGAAGACCAACTACCGGAGTACATGAGGCTTACGATCTTGGAGCTC 3840
DB 15652 CGGCACTTCTTGAAGACCAACTACCGGAGTACATGAGGCTTACGATCTTGGAGCTC 15711
QY 3841 GATCGCTTCCGCTGACGCTCCACGCGGAGTTCATGAGGCTTCCGCGGAGG 3900
DB 15712 GATCGCTTCCGCTGACGCTCCACGCGGAGTTCATGAGGCTTCCGCGGAGG 15771
QY 3901 GATACCTGCTGCGCGGCTTCCGCGGAGTTCATGAGGCTTCCGCGGAGG 3960
DB 15772 GATACCTGCTGCGCGGCTTCCGCGGAGTTCATGAGGCTTCCGCGGAGG 15831
QY 3961 CTGCTGCGGCTGACGCGGAGTTCGCGGCTGAGAGGCTTCCGCGGAGG 4020
DB 15832 CTGCTGCGGCTGACGCGGAGTTCGCGGCTGAGAGGCTTCCGCGGAGG 15891
QY 4021 GTGATCTTGGTGCAGATCGATTCAATCGTTGCGATGAGAGCTTTCAGAGAGG 4080
DB 15892 GTGATCTTGGTGCAGATCGATTCAATCGTTGCGATGAGAGCTTTCAGAGAGG 15951
QY 4081 CTGATAGGAGATCGCATCACGAGTTCGAGTACCGGAACTTCGCGCTCGTGGCG 4140
DB 15952 CTGATAGGAGATCGCATCACGAGTTCGAGTACCGGAACTTCGCGCTCGTGGCG 16011
QY 4141 TCCGCTTTCGCGGAGCTGAGAGTTCATGAGGCGGAGGCGGAGGAGGAGG 4200
DB 16012 TCCGCTTTCGCGGAGCTGAGAGTTCATGAGGCGGAGGCGGAGGAGGAGG 16071
QY 4201 GAGGTTGCGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4233
DB 16072 GAGGTTGCGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 16104

RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 635457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Deyon
APPLICANT: Goetlich, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,969
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 100.0%; Score 4233; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGATCAATCAGCTTCTGAACGAGCTTCAGACACCAAGGTATCAAGCTGCGCGCAT 60
DB 11872 ATGACGATCAATCAGCTTCTGAACGAGCTTCAGACACCAAGGTATCAAGCTGCGCGCAT 11931
QY 61 GGGGAGCGCTCCAGATTAAGGCCCCCAAGACCCCTGAAACCCGAACTGCTGCTCGA 120
DB 11932 GGGGAGCGCTCCAGATTAAGGCCCCCAAGACCCCTGAAACCCGAACTGCTGCTCGA 11991
QY 121 ATCTCGAGACAAAGACAGATCTGAGATGCTCCGTAGAGATCTCCCGGAGATCC 180
DB 11992 ATCTCGAGACAAAGACAGATCTGAGATGCTCCGTAGAGATCTCCCGGAGATCC 12051
QY 181 ATCTGCCCCCGCCAGCCGAGCGGACGCTCGCTTCTCTCAAGACATCCAAAGATCC 240
DB 12052 ATCTGCCCCCGCCAGCCGAGCGGACGCTCGCTTCTCTCAAGACATCCAAAGATCC 12111
QY 241 TACTGCTGGCGCGGACAGAGCGTTTAAAGTCCCAAGGCGGATTCACAGCTTATCGCGAA 300
DB 12112 TACTGCTGGCGCGGACAGAGCGTTTAAAGTCCCAAGGCGGATTCACAGCTTATCGCGAA 12171
QY 301 TACGACTGATCGATCTCGACGTGCGGAGCTGAGCGCGCTTTCGAAAGTCTGCGG 360
DB 12172 TACGACTGATCGATCTCGACGTGCGGAGCTGAGCGCGCTTTCGAAAGTCTGCGG 12231
QY 361 CGGACGACATGCTTGGGCGCCACACGCTGCGGACATGATGAGGTATCGAGCTTAA 420
DB 12232 CGGACGACATGCTTGGGCGCCACACGCTGCGGACATGATGAGGTATCGAGCTTAA 12291
QY 421 GTTCGACCGGACATTCGAGATCATGATTCGCGGCGCTGACCGGAGACACAGGGAAGG 480
DB 12292 GTTCGACCGGACATTCGAGATCATGATTCGCGGCGCTGACCGGAGACACAGGGAAGG 12351
QY 481 AGGCTGCTGCTGTTGGAGATGCGATGTCGACCGATCTATGACACCGAGCGCCCTCG 540
DB 12352 AGGCTGCTGCTGTTGGAGATGCGATGTCGACCGATCTATGACACCGAGCGCCCTCG 12411
QY 541 CTCTATCAGCTGCTGCGCGCTTGGCTGAGACGAGCGGAAACCCGCTGCTGCTCACTATC 600
DB 12412 CTCTATCAGCTGCTGCGCGCTTGGCTGAGACGAGCGGAAACCCGCTGCTGCTCACTATC 12471
QY 601 GATCTCATTAACCTTAAACCTTAAAGGACGCTGTCATATCTTAAAGACCTGCTCACTTC 660
DB 12472 GATCTCATTAACCTTAAACCTTAAAGGACGCTGTCATATCTTAAAGACCTGCTCACTTC 12531
QY 661 TACGAGATCCGAGACCTTCTCCCTGCTGAGAGCTCGTACCGGAGTATGATCTC 720
DB 12532 TACGAGATCCGAGACCTTCTCCCTGCTGAGAGCTCGTACCGGAGTATGATCTC 12591
QY 721 GCGCTGAGCTCTGCAAGAGTCTGAGGCGCATCAACGATGATGATTACTGAAACCG 780
DB 12592 GCGCTGAGCTCTGCAAGAGTCTGAGGCGCATCAACGATGATGATTACTGAAACCG 12651

QY 781 GGCATGCGGAGCTCCACCTCCCGGACGCTTCGATGAAGGCGCATCATCACTC 840
DB 12652 GGCATGCGGAGCTCCACCTCCCGGACGCTTCGATGAAGGCGCATCATCACTC 12711
QY 841 AAGGATCCGCTTCGAGACAGGAGGATGAGTCCGCTGAGATCTCTGGGCTGATG 900
DB 12712 AAGGATCCGCTTCGAGACAGGAGGATGAGTCCGCTGAGATCTCTGGGCTGATG 12771
QY 901 AAGCGCGTGTGGGAGACCGGCGTGAACCCGAGCGGCTCATCTGCTGATTTTCC 960
DB 12772 AAGCGCGTGTGGGAGACCGGCGTGAACCCGAGCGGCTCATCTGCTGATTTTCC 12831
QY 961 GAGGTATCGGCGCTGAGACGCGCCCGGATTAAGCTCAACATACGCTTCAAC 1020
DB 12832 GAGGTATCGGCGCTGAGACGCGCCCGGATTAAGCTCAACATACGCTTCAAC 12891
QY 1021 CGGCTCCCGCTCATCCGCGCTGAAAGATATCAACGAGGACCTTACGCTGATGCTC 1080
DB 12892 CGGCTCCCGCTCATCCGCGCTGAAAGATATCAACGAGGACCTTACGCTGATGCTC 12951
QY 1081 CTGACATCGACACCACTCGGACAAAGCTTCGAAACGCGGCTTAAGCTTCAAG 1140
DB 12952 CTGACATCGACACCACTCGGACAAAGCTTCGAAACGCGGCTTAAGCTTCAAG 13011
QY 1141 CAGCTGAGGAGAGCATGATCATCTGACGACGATTAAGCTTCAAGCTCAGCGAGGCC 1200
DB 13012 CAGCTGAGGAGAGCATGATCATCTGACGACGATTAAGCTTCAAGCTCAGCGAGGCC 13071
QY 1201 GCCCGGCTCTGGGAGTCCAAAGAGCGCATTTGTTCCCGTGTGCTACAGAGCGGCTT 1260
DB 13072 GCCCGGCTCTGGGAGTCCAAAGAGCGCATTTGTTCCCGTGTGCTACAGAGCGGCTT 13131
QY 1261 AACCGCAAGTGTGTGTCTACCTCTGTCAGAGAGCTTCGAACTCCGCTGTACACG 1320
DB 13132 AACCGCAAGTGTGTGTCTACCTCTGTCAGAGAGCTTCGAACTCCGCTGTACACG 13191
QY 1321 ACGCAATCTCTCAGCTGCTGATCATCACTCTACAGACGATGAGGACCTCTGC 1380
DB 13192 ACGCAATCTCTCAGCTGCTGATCATCACTCTACAGACGATGAGGACCTCTGC 13251
QY 1381 CTGCGTGGGACATCTGTCAGAGAGTGTCCGCGCCGACCTTCTGACGACATGCTCA 1440
DB 13252 CTGCGTGGGACATCTGTCAGAGAGTGTCCGCGCCGACCTTCTGACGACATGCTCA 13311
QY 1441 GGTACGTGTTTTTCTCGGCGGCTCACTAGAGAACATGGGAGTGAACAGGTGCGCT 1500
DB 13312 GGTACGTGTTTTTCTCGGCGGCTCACTAGAGAACATGGGAGTGAACAGGTGCGCT 13371
QY 1501 TCGCTTCGCTGCGACGATGAAGCGCGGAGCGGAAACGCGACCAACGCGTGTG 1560
DB 13372 TCGCTTCGCTGCGACGATGAAGCGCGGAGCGGAAACGCGACCAACGCGTGTG 13431
QY 1561 ACGAGCATACGCTGACCGCTGTTCCGCGCGGCTGAGACGCTGCCATGCACTC 1620
DB 13432 ACGAGCATACGCTGACCGCTGTTCCGCGCGGCTGAGACGCTGCCATGCACTC 13491
QY 1621 GCCGTGTGTGCGGCGGCAAGACGCTCACTGAGAGAGCTTCCGCGCTTCCGCGG 1680
DB 13492 GCCGTGTGTGCGGCGGCAAGACGCTCACTGAGAGAGCTTCCGCGCTTCCGCGG 13551
QY 1681 CTGAGCGCGCTGCGGACGAGGAGGCGCGCGGAAACATTTGCTGCGGTGTGATG 1740
DB 13552 CTGAGCGCGCTGCGGACGAGGAGGCGCGCGGAAACATTTGCTGCGGTGTGATG 13611
QY 1741 GAGAAAGCTGAGGACAGGTTGTGCGGTTCTGCGGTGCTGAGTCAAGCGCGCTTAC 1800
DB 13612 GAGAAAGCTGAGGACAGGTTGTGCGGTTCTGCGGTGCTGAGTCAAGCGCGCTTAC 13671
QY 1801 GTGCGATGATGCGGACCTTACCGGCGAGCGCTTACCTACTCTCTGATATGATG 1860
DB 13672 GTGCGATGATGCGGACCTTACCGGCGAGCGCTTACCTACTCTCTGATATGATG 13731
QY 1861 GTAAAGCTGTGCTGACGAGCATGTGATGAGCAACTGTCAATGCGCGCGGAGTCC 1920

Db 13732 GTAAAGCTCGTGTGACCCAGCATGGCTGATGCAAACTGTCAATGACCCGCGGGATC 13791
Qy 1991 CAGGGGCTGCTGTGAGGAGGCGGCGTCCGAAAGCGACGCGACCAAGCTCCGATGANG 1980
Db 13792 CAGCGGCTGTGTGTAGGAGGCGCGCTCGAAGGCGACGCGACCAAGCTCCGATGANG 13851
Qy 1981 CCCATTGAGACACTTCGGATCTCGGATATGTCATCTACACTTCGGGATCCAGAGTTG 2040
Db 13852 CCCATTGAGACACTTCGGATCTCGGATATGTCATCTACACTTCGGGATCCAGAGTTG 13911
Qy 2041 CCCAAGGGGGTGTATGATGATCATTCGGGGTGCCTCAACCACTCTGGACATCAACGAG 2100
Db 13912 CCCAAGGGGGTGTATGATGATCATTCGGGGTGCCTCAACCACTCTGGACATCAACGAG 13971
Qy 2101 CGCTTCGAATATGGGCGCGGAGACAGGGGCTGGGGCTCTCCTCGCTGAGCTTCGATCTC 2160
Db 13972 CGCTTCGAATATGGGCGCGGAGACAGGGGCTGGGGCTCTCCTCGCTGAGCTTCGATCTC 14031
Qy 2161 TCGCTCTATGATGATGTTGGGATCCTGGCGGCGGCGGATCGATCGTGGTGCAGACCG 2220
Db 14032 TCGCTCTATGATGATGTTGGGATCCTGGCGGCGGCGGATCGATCGTGGTGCAGACCG 14091
Qy 2221 TCCAAAGCTGCGGATCCGCGCATTTGGGCAAGTTGATCGAAAGAGAGGTGACGGTG 2280
Db 14092 TCCAAAGCTGCGGATCCGCGCATTTGGGCAAGTTGATCGAAAGAGAGGTGACGGTG 14151
Qy 2281 TGGAACTGGTGGCCGGCGCTGATGGGATGCTCGCTCGAGACATTTGAGGGTCCCGCAT 2340
Db 14152 TGGAACTGGTGGCCGGCGCTGATGGGATGCTCGCTCGAGACATTTGAGGGTCCCGCAT 14211
Qy 2341 TCGCTCGTAGAGTCTGTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGTGGGCGTG 2400
Db 14212 TCGCTCGTAGAGTCTGTGCGGCTTTCGCTGCTGAGCGGCGACTGGATCCCGTGGGCGTG 14271
Qy 2401 CCTGGCGAGCTTCAGAGCCATCAGGCCCGGCGTGTGCTGATCAGCTTGGCGGGGCGCAC 2460
Db 14272 CCTGGCGAGCTTCAGAGCCATCAGGCCCGGCGTGTGCTGATCAGCTTGGCGGGGCGCAC 14331
Qy 2461 GAAAGGTGATCTGTGTCATCGGGATACCCCGTGAAGAAAGTGAACCTTATCGTGGCGAGC 2520
Db 14332 GAAAGGTGATCTGTGTCATCGGGATACCCCGTGAAGAAAGTGAACCTTATCGTGGCGAGC 14391
Qy 2521 ATCCCTACGGGCGCTCGCTGCGCAACGAGACTTCTCAAGTCTCGATGAGGCGCTCGAA 2580
Db 14392 ATCCCTACGGGCGCTCGCTGCGCAACGAGACTTCTCAAGTCTCGATGAGGCGCTCGAA 14451
Qy 2581 CCGCGCGCGGTGCGGTTCCGGGAGCACTTACATTTGCGGGGTGCGGCTGCGACTGGGC 2640
Db 14452 CCGCGCGCGGTGCGGTTCCGGGAGCACTTACATTTGCGGGGTGCGGCTGCGACTGGGC 14511
Qy 2641 TACTGGCCCGATGAGAGAAAGCGCGCAAGACTTCTCTGTGACACCCGAGACCGGGAG 2700
Db 14512 TACTGGCCCGATGAGAGAAAGCGCGCAAGACTTCTCTGTGACACCCGAGACCGGGAG 14571
Qy 2701 CGCCTCTACAGACCGGCGATCTGGGCGGCTACCTGCGCGATGAAACATCGAGTTCAATG 2760
Db 14572 CGCCTCTACAGACCGGCGATCTGGGCGGCTACCTGCGCGATGAAACATCGAGTTCAATG 14631
Qy 2761 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATACCGGTTGAGCTCGGGGAAATCGAG 2820
Db 14632 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATACCGGTTGAGCTCGGGGAAATCGAG 14691
Qy 2821 GAAAGCTCAAGTCCGATCCGAAAGTACCGGACGCGGTGATGTTGTCCTCGGGAACGAC 2880
Db 14692 GAAAGCTCAAGTCCGATCCGAAAGTACCGGACGCGGTGATGTTGTCCTCGGGAACGAC 14751
Qy 2881 GCGCGGAACAAGCTCTTCTAGCTATGTCCTCGGAGGCGACACGAGAGACGCGCTGCC 2940
Db 14752 GCGCGGAACAAGCTCTTCTAGCTATGTCCTCGGAGGCGACACGAGAGACGCGCTGCC 14811
Qy 2941 GAGCAGACGCGAGCTTCAAGCCGAGCGGATCGACGCGAGAGCAACGCGCGCGAAGCG 3000

Db 14812 GAGCAGACGCGAGCTTCAAGCCGAGCGGATCGACCGGAGAGCAACGCGCGCGAAGCG 14871
Qy 3001 GACGCGCTTGAAGCGAGCGGAGAGGATGATTAAGCTCGCTCGACACGAGACTCCGAGG 3060
Db 14872 GACGCGCTTGAAGCGAGCGGAGAGGATGATTAAGCTCGCTCGACACGAGACTCCGAGG 14931
Qy 3061 GACCTGAGCGGAAAGCCGCTGTCGATCTGAACGCGGACGAGATCCGCGGAGCGGGGCTG 3120
Db 14932 GACCTGAGCGGAAAGCCGCTGTCGATCTGAACGCGGACGAGATCCGCGGAGCGGGGCTG 14991
Qy 3121 GACCTGAGCGGCGCTCGCGTGAAGGTCGAAAGCTTCTTGAAGGCGCCGATTCCTTTGTT 3180
Db 14992 GACCTGAGCGGCGCTCGCGTGAAGGTCGAAAGCTTCTTGAAGGCGCCGATTCCTTTGTT 15051
Qy 3181 GAGTTGGTGCATTCCTGAGCTGCTTGAAGCGAGCGGAGCGCGACCGGCGCGACCTTCC 3240
Db 15052 GAGTTGGTGCATTCCTGAGCTGCTTGAAGCGAGCGGAGCGCGACCGGCGCGACCTTCC 15111
Qy 3241 AAATTCGTTATCCATCGGCGGGGAGCAAGTACCCTGCAAACTTACGCTATGTCAAA 3300
Db 15112 AAATTCGTTATCCATCGGCGGGGAGCAAGTACCCTGCAAACTTACGCTATGTCAAA 15171
Qy 3301 TCCGCGCGCATCGAGGGCGTGAAGAGGCTTCTTATTTACACCCGTTGAAGACCGT 3360
Db 15172 TCCGCGCGCATCGAGGGCGTGAAGAGGCTTCTTATTTACACCCGTTGAAGACCGT 15231
Qy 3361 TTGCTGAAGCTCTCGATCAAGGGATCGAGGCGGAGGCGAGCTTCCGCAAACTTCGAC 3420
Db 15232 TTGCTGAAGCTCTCGATCAAGGGATCGAGGCGGAGGCGAGCTTCCGCAAACTTCGAC 15291
Qy 3421 GTGTTGATGAAGCGGCGTTTCAACTCTGTTCTGTGGCAGAGATCGACCGCATCGAGTCG 3480
Db 15292 GTGTTGATGAAGCGGCGTTTCAACTCTGTTCTGTGGCAGAGATCGACCGCATCGAGTCG 15351
Qy 3481 CTGTATGATGCTGTGTGCGAGAAATTTGCTGCTGAGGCGCGGATATATGCGCAGCTC 3540
Db 15352 CTGTATGATGCTGTGTGCGAGAAATTTGCTGCTGAGGCGCGGATATATGCGCAGCTC 15411
Qy 3541 CTGATGAGAGAGGCGCTCTCGAACAATGGCGGCTGTGCGGTGGGCGAAATTCATTTT 3600
Db 15412 CTGATGAGAGAGGCGCTCTCGAACAATGGCGGCTGTGCGGTGGGCGAAATTCATTTT 15471
Qy 3601 GAAAGGTTCCGCGCGGTTCTCGACCTGCGACATTTCGAGCTTTTACGTCGACGCGATGCTG 3660
Db 15472 GAAAGGTTCCGCGCGGTTCTCGACCTGCGACATTTCGAGCTTTTACGTCGACGCGATGCTG 15531
Qy 3661 GCGCGGCGGATGAGACCGCGGCAAGTTTCAAGTCTGTACGCTCGGTCAGAGATTCTCACCG 3720
Db 15532 GCGCGGCGGATGAGACCGCGGCAAGTTTCAAGTCTGTACGCTCGGTCAGAGATTCTCACCG 15591
Qy 3721 AGGCGCGCAGAGCGCGGCGCCCTCCGCGCGCGAGCACTTGGCGGATATGCTT 3780
Db 15592 AGGCGCGCAGAGCGCGGCGCCCTCCGCGCGCGAGCACTTGGCGGATATGCTT 15651
Qy 3781 CGGCACTTCTTGAGGAGCAACCTACCGGATGATGTCCTTACAGTCTTGTGAGGCTC 3840
Db 15652 CGGCACTTCTTGAGGAGCAACCTACCGGATGATGTCCTTACAGTCTTGTGAGGCTC 15711
Qy 3841 GATGCTTGCCTGACGCTCCACCGGCAAGGTCGATGTAAGGCTTCGCGAGCGGAG 3900
Db 15712 GATGCTTGCCTGACGCTCCACCGGCAAGGTCGATGTAAGGCTTCGCGAGCGGAG 15771
Qy 3901 GATGCTTGCCTGACGCTCCACCGGCAAGGTCGATGTAAGGCTTCGCGAGCGGAG 3960
Db 15772 GATGCTTGCCTGACGCTCCACCGGCAAGGTCGATGTAAGGCTTCGCGAGCGGAG 15831
Qy 3961 CTGCTGCGGCTCTGATCGGAGGTCCTCGGCTGAGAGTGTCTCGGCTTCACAGAGCTTC 4020
Db 15832 CTGCTGCGGCTCTGATCGGAGGTCCTCGGCTGAGAGTGTCTCGGCTTCACAGAGCTTC 15891
Qy 4021 GTCGATCTTGGTGGACATCGATTCAATGTTGCGATGAGAGGCTTGTGCAAGAGG 4080
Db 15892 GTCGATCTTGGTGGACATCGATTCAATGTTGCGATGAGAGGCTTGTGCAAGAGG 15951

QY 4081 CTGGATAGGAGATCCGCATCAACCGAGTTGTTCCAGTACCAGAACTCCGCTCGTGGG 4140
DB 15952 CTGGATAGGAGATCCGCATCAACCGAGTTGTTCCAGTACCAGAACTCCGCTCGTGGG 16011
QY 4141 TCCGATTTCCGCGGAGAGCTCGAGAGATCTAGATCAGCGGCCGAACATGACGACGAGT 4200
DB 16012 TCCGATTTCCGCGGAGAGCTCGAGAGATCTAGATCAGCGGCCGAACATGACGAGT 16071
QY 4201 GAGGTTCCGCGCAAGGCGAGAGAGCTAGCTAA 4233
DB 16072 GAGGTTCCGCGCAAGGCGAGAGAGCTAGCTAA 16104

RESULT 4
US-09-568-480-1
Sequence 1, Application US/09568480
Patent No. 6355458
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Rose
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,480
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-568-480-1

Query Match 100.0%; Score 4233; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACGATCAATAGCTTCTGAAAGAGCTCGAGACACAGGTTTCAAGCTGGGGCGCAT 60
DB 11872 ATGACGATCAATAGCTTCTGAAAGAGCTCGAGACACAGGTTTCAAGCTGGGGCGCAT 11931
QY 61 GGGAGAGCGCTCAGATTAAGAGCCGCCAAGAACGCCCTGAAACCGAAGCTGCTGCTCA 120
DB 11932 GGGAGAGCGCTCAGATTAAGAGCCGCCAAGAACGCCCTGAAACCGAAGCTGCTGCTCA 11991
QY 121 ATCTCCGAGCAAAAGACAGATCTGAGAGCTCTCCGTCAGAGACTCCCGCGAATCC 180
DB 11992 ATCTCCGAGCAAAAGACAGATCTGAGAGCTCTCCGTCAGAGACTCCCGCGAATCC 12051
QY 181 ATGTGCGCGCCCGAGCGGAGCGCTCCGTTCTCTCAGACATCCAAAGATCC 240
DB 12052 ATGTGCGCGCCCGAGCGGAGCGCTCCGTTCTCTCAGACATCCAAAGATCC 12111
QY 241 TACTGCTGGGCGGAGCAGAGCGTTTACGATCCCGAGCGAGTCCACGCTTATCCGAA 300
DB 12112 TACTGCTGGGCGGAGCAGAGCGTTTACGATCCCGAGCGAGTCCACGCTTATCCGAA 12171
QY 301 TACGACTGTACGAGATCTCGAGTGCAGAGCTGAGCGCGCTTTGAGAAAGTGTGCGG 360
DB 12172 TACGACTGTACGAGATCTCGAGTGCAGAGCTGAGCGCGCTTTGAGAAAGTGTGCGG 12231
QY 361 CGGCAAGATGTTTGGGGCGGACAGCTCCCGAATGATGAGGTGATTCGAGCTTAA 420
DB 12232 CGGCAAGATGTTTGGGGCGGACAGCTCCCGAATGATGAGGTGATTCGAGCTTAA 12291
QY 421 GTTCGACGCGGACATCGAGATCATGATCTGCGGGGCTCGACCGGAGCACACGGAGAGG 480

DB 12292 GTTCGACGCGGACATCGAGATCATGATCTGCGGGGCTGACCGGAGCACACGGAGAGG 12351
QY 481 AGGCTGTGTGTTGCGAGATTCGATGTGCAACCGCATCTATGACACCGAGGCCCTCCG 540
DB 12352 AGGCTGTGTGTTGCGAGATTCGATGTGCAACCGCATCTATGACACCGAGGCCCTCCG 12411
QY 541 CTCTATCAAGTGTGCGCGGTTGAGGAGCGGCAAAACCGCTGTGCTGCTAGTATC 600
DB 12412 CTCTATCAAGTGTGCGCGGTTGAGGAGCGGCAAAACCGCTGTGCTGCTAGTATC 12471
QY 601 GATCTCATTAAGTTGACCTAGGAGCGCTGTCCATCATCTTCAAGAGCTGAGCTTAC 660
DB 12472 GATCTCATTAAGTTGACCTAGGAGCGCTGTCCATCATCTTCAAGAGCTGAGCTTAC 12531
QY 661 TACGAAGATCCGAGACCTCTCTCCCTGTCTCTGAGAGCTCTGTTACCGGATTAATGATC 720
DB 12532 TACGAAGATCCGAGACCTCTCTCCCTGTCTCTGAGAGCTCTGTTACCGGATTAATGATC 12591
QY 721 GCGCTGAGGTCGCAAGAGCTGAGGCGCATCAACGATCGATGATTAATGAGAGCGG 780
DB 12592 GCGCTGAGGTCGCAAGAGCTGAGGCGCATCAACGATCGATGATTAATGAGAGCGG 12651
QY 781 CGCATCGCGAGCTCCACCTCCGCGAGCGCTTCCGATGAGAGCGGATCGATCTACCTG 840
DB 12652 CGCATCGCGAGCTCCACCTCCGCGAGCGCTTCCGATGAGAGCGGATCGATCTACCTG 12711
QY 841 AAGGAGATCCGCTTCCGCGACCGAGCAATGGCTCCGCTCGGACTCCGCGGGTGGATTG 900
DB 12712 AAGGAGATCCGCTTCCGCGACCGAGCAATGGCTCCGCTCGGACTCCGCGGGTGGATTG 12771
QY 901 AAGCGCGGTCGCGGAGAGCGGAGCTGACCCCGAGCGGCTGATCTGAGCTCATTTTCC 960
DB 12772 AAGCGCGGTCGCGGAGAGCGGAGCTGACCCCGAGCGGCTGATCTGAGCTCATTTTCC 12831
QY 961 GAGGTATCGGCGCTGAGAGCGGAGCGGCTTACGCTCAATTAACGCTTTCAAC 1020
DB 12832 GAGGTATCGGCGCTGAGAGCGGAGCGGCTTACGCTCAATTAACGCTTTCAAC 12891
QY 1021 CGGCTCCCGTCATTCGCGGCTGAAAGATATCACCGGGGATTTCAAGCTGATGCTTCC 1080
DB 12892 CGGCTCCCGTCATTCGCGGCTGAAAGATATCACCGGGGATTTCAAGCTGATGCTTCC 12951
QY 1081 CTGGAATCGACACCACTGCGAGCAAGAGCTTCAAGAGCGGCTTAAGGTTTAAAGAG 1140
DB 12952 CTGGAATCGACACCACTGCGAGCAAGAGCTTCAAGAGCGGCTTAAGGTTTAAAGAG 13011
QY 1141 CAGCTGTGGAGAGATGATGATCACTGCAAGCTTAAGCGGTATCGAGTCCAGAGAGCC 1200
DB 13012 CAGCTGTGGAGAGATGATGATCACTGCAAGCTTAAGCGGTATCGAGTCCAGAGAGCC 13071
QY 1201 GCGCGGTCGTTGGGATCCAGAGCGGATTTCCCGTGTGCTCAAGAGCGGCTT 1260
DB 13072 GCGCGGTCGTTGGGATCCAGAGCGGATTTCCCGTGTGCTCAAGAGCGGCTT 13131
QY 1261 AACGAGCAAGTGTGTGCTCACTGTTGAGAGGCTTGAAGCTCCGTTGTACACAGC 1320
DB 13132 AACGAGCAAGTGTGTGCTCACTGTTGAGAGGCTTGAAGCTCCGTTGTACACAGC 13191
QY 1321 ACGGAGACTCTCAAGCTGTGATCACTGCTTACGAGACGATGGGAGCTCGTC 1380
DB 13192 ACGGAGACTCTCAAGCTGTGATCACTGCTTACGAGACGATGGGAGCTCGTC 13251
QY 1381 CTGCGTGGGAGATGATGATGAGAGGTTCCGCGGACCTTGTGAGAGCATGTGCGAA 1440
DB 13252 CTGCGTGGGAGATGATGATGAGAGGTTCCGCGGACCTTGTGAGAGCATGTGCGAA 13311
QY 1441 GGTACGTGTTTCTCCGCGGCTCACTGAGAGCACTAGGGGTGAACAGGTGCGCTGT 1500
DB 13312 GGTACGTGTTTCTCCGCGGCTCACTGAGAGCACTAGGGGTGAACAGGTGCGCTGT 13371
QY 1501 TCGCTTCGCGCTGCGGAGTGAAGCGCGGAGCGCAACCGCATCCGCTGCTG 1560
DB 13372 TCGCTTCGCGCTGCGGAGTGAAGCGCGGAGCGCAACCGCATCCGCTGCTG 13431

1561 AGCGACATACGCTGACAGGCTGTTTCGCGCGCGGCTGAGCACTGCGCATGAGCTC 1620
13432 AGCGACATACGCTGACAGGCTGTTTCGCGCGCGGCTGAGCACTGCGCATGAGCTC 13491
1621 GCCCGGTGTGCGCGCGCAAGCGCTCACATGCAAGAGCTTTCCGCGCGTTCGCGCGCA 1680
13492 GCCCGGTGTGCGCGCGCAAGCGCTCACATGCAAGAGCTTTCCGCGCGTTCGCGCGCA 13551
1681 CTTGCGCGCGGCTGCGGAGCGAGGCGGAGCGCGGAAACATGTTGCGCGGTGATG 1740
13552 CTTGCGCGCGGCTGCGGAGCGAGGCGGAGCGCGGAAACATGTTGCGCGGTGATG 13611
1741 GAGAAAGCTGCGGAGCAAGTTGTCGCGGTCTCGCGGTCTGAGTCAAGCGCGCTAC 1800
13612 GAGAAAGCTGCGGAGCAAGTTGTCGCGGTCTCGCGGTCTGAGTCAAGCGCGCTAC 13671
1801 GTGCGGATCGATCCGACCTTACCGGCGGAGCGATTCACATCTTCCTGATCATGTAG 1860
13672 GTGCGGATCGATCCGACCTTACCGGCGGAGCGATTCACATCTTCCTGATCATGTAG 13731
1861 GTAAAGCTGTGCTGACGAGCGAGCTGAGTGGCAATGTCATGCGCGCGGAGTAC 1920
13732 GTAAAGCTGTGCTGACGAGCGAGCTGAGTGGCAATGTCATGCGCGCGGAGTAC 13791
1921 CAGCGGCTGTGCTGAGCGAGCGCGGCTGAGCGAGCGAGCGAGCGAGCTTCGATGATG 1980
13792 CAGCGGCTGTGCTGAGCGAGCGCGGCTGAGCGAGCGAGCGAGCGAGCTTCGATGATG 13851
1981 CCCATTCAAGACCTTTCGATCTGCGATCTGCTGATCTCACTCACTTCGCGGATCCAGAGTTC 2040
13852 CCCATTCAAGACCTTTCGATCTGCGATCTGCTGATCTCACTCACTTCGCGGATCCAGAGTTC 13911
2041 CCCAAGGGGGTGAATGATGATGATGAGGGTGGCGTCAACATCTGGAATGCAAGAG 2100
13912 CCCAAGGGGGTGAATGATGATGAGGGTGGCGTCAACATCTGGAATGCAAGAG 13971
2101 CGCTTCGAAATAGAGCGCGGAGACAGGGTGTGCGCTCTCTCGCTGAGTTCGATCTC 2160
13972 CGCTTCGAAATAGAGCGCGGAGACAGGGTGTGCGCTCTCTCGCTGAGTTCGATCTC 14011
2161 TCGGCTATGATGTGTTGGGATCTCTGCGGCGGCGGAGTGCATGTTGCGGAGCGG 2220
14032 TCGGCTATGATGTGTTGGGATCTCTGCGGCGGCGGAGTGCATGTTGCGGAGCGG 14091
2221 TCCAGGCTGCGGATCCGCGCGATGAGGAGGATGATGCAACGAGAGAGAGTGAAGCTG 2280
14092 TCCAGGCTGCGGATCCGCGCGATGAGGAGGATGATGCAACGAGAGAGAGTGAAGCTG 14151
2281 TGGAACTGAGTCCGCGCGTGAATGCGATGCTCGTCAAGCATTTTGAAGGATGAGCCCGAT 2340
14152 TGGAACTGAGTCCGCGCGTGAATGCGATGCTCGTCAAGCATTTTGAAGGATGAGCCCGAT 14211
2341 TCGCTCGTGAAGTCTCTGCGGCTTTTCGCTGATGAGCGGAGCTGATCCCGTGGAGCTG 2400
14212 TCGCTCGTGAAGTCTCTGCGGCTTTTCGCTGATGAGCGGAGCTGATCCCGTGGAGCTG 14271
2401 CCTGCGAGCTTCCAGGCGATCAAGCGCGCGGCTGTGAGTGAAGCTGAGGCGGCGCAC 2460
14272 CCTGCGAGCTTCCAGGCGATCAAGCGCGCGGCTGTGAGTGAAGCTGAGGCGGCGCAC 14331
2461 GAAAGCTGATCTGCTCAATCGAGTACCGGAGAGAACTGCACTATCTGAGGCGAGC 2520
14332 GAAAGCTGATCTGCTCAATCGAGTACCGGAGAGAACTGCACTATCTGAGGCGAGC 14391
2521 ATCCCTTACGCGCGTCCGCTGCGCAACGAGCTTTCACATGTCGATGAGGCGCTGCA 2580
14392 ATCCCTTACGCGCGTCCGCTGCGCAACGAGCTTTCACATGTCGATGAGGCGCTGCA 14451
2581 CCGCGCGGCTTCTGAGGTTCCGCGGAGCACTTCAATGCGGAGTGGCGTGGCACTGAGC 2640
14452 CCGCGCGGCTTCTGAGGTTCCGCGGAGCACTTCAATGCGGAGTGGCGTGGCACTGAGC 14511

2641 TACTGGCGGATGAAAGAGACGCGCAAGAGCTTCTCTGTCACCCCGAGACCGGGAG 2700
14512 TACTGGCGGATGAAAGAGACGCGCAAGAGCTTCTCTGTCACCCCGAGACCGGGAG 14571
2701 GCCCTTCAAGACCGGAGTCTGGGCGGCTACCTGCGCGGATGGAATACTGAGTTCTAG 2760
14572 GCCCTTCAAGACCGGAGTCTGGGCGGCTACCTGCGCGGATGGAATACTGAGTTCTAG 14631
2761 GGGCGTAGAGCAACCAAACTCAAGCTTTCGCGGATACCGGTTGAGCTCGGGGAAATCGAG 2820
14632 GGGCGTAGAGCAACCAAACTCAAGCTTTCGCGGATACCGGTTGAGCTCGGGGAAATCGAG 14691
2821 GAAACGCTCAAGTGCATCCGACGTAACGCGAGCGGATGATGTCGCCGTGCGGACGAC 2880
14692 GAAACGCTCAAGTGCATCCGACGTAACGCGAGCGGATGATGTCGCCGTGCGGACGAC 14751
2881 GCGGCGAACAAGCTCTTCTAGCTTATGCTTATGCTTCCCGAGGAGCAACGAGACCGCTCC 2940
14752 GCGGCGAACAAGCTCTTCTAGCTTATGCTTATGCTTCCCGAGGAGCAACGAGACCGCTCC 14811
2941 GAGCGAGACGCGAGCTCAAGACCGAGGAGTTCGACGCGAGAGCAACGCGCGGAGCG 3000
14812 GAGCGAGACGCGAGCTCAAGACCGAGGAGTTCGACGCGAGAGCAACGCGCGGAGCG 14871
3001 GACGCTTGAACGACGCGGAGAGGAGTTCGACGCTGCTGACACGAGCTTCGAGAG 3060
14872 GACGCTTGAACGACGCGGAGAGGAGTTCGACGCTGCTGACACGAGCTTCGAGAG 14931
3061 GACCTGACGAGAAAGCGGCTGTGATTTGACCGGAGAGATTCGCGGAGAGCGGGCTG 3120
14932 GACCTGACGAGAAAGCGGCTGTGATTTGACCGGAGAGATTCGCGGAGAGCGGGCTG 14991
3121 GACGCTACGCGGCGGCGGATGAGCTGCGGAGCTTCTGAGGCGCGGATTCGCTTGT 3180
14992 GACGCTACGCGGCGGCGGATGAGCTGCGGAGCTTCTGAGGCGCGGATTCGCTTGT 15051
3181 GAGTGTGATGATTCCTGAGCTGCTTGAAGAGCGTGAAGCGGAGCGGAGCGGAGCTTC 3240
15052 GAGTGTGATGATTCCTGAGCTGCTTGAAGAGCGTGAAGCGGAGCGGAGCGGAGCTTC 15111
3241 AAATTCGTTTCCATCGCGCGGAGAGCGATCCCGGTGCAACCTTACGCTATGTCAAA 3300
15112 AAATTCGTTTCCATCGCGCGGAGAGCGATCCCGGTGCAACCTTACGCTATGTCAAA 15171
3301 TCCGCGGATCGAGGAGGAGTGAAGAGGAGCTTCTATTTACACCCGTTGAGGAGCGT 3360
15172 TCCGCGGATCGAGGAGGAGTGAAGAGGAGCTTCTATTTACACCCGTTGAGGAGCGT 15231
3361 TTGCTGAAGCTCTCGATCAAGGATCGAGCGGAGCGAGCGATTCGCAAAACTTCGAC 3420
15232 TTGCTGAAGCTCTCGATCAAGGATCGAGCGGAGCGAGCGATTCGCAAAACTTCGAC 15291
3421 GTGTTGATGAGGCGGCTTCAACCTCTGTTGAGGAGAGATGAGCGCATGAGTGC 3480
15292 GTGTTGATGAGGCGGCTTCAACCTCTGTTGAGGAGAGATGAGCGCATGAGTGC 15351
3481 CTGATGATGATGCTGCGGAGAAATTTGCTGCTGAGGAGCGGATATATGCGGAGCTC 3540
15352 CTGATGATGATGCTGCGGAGAAATTTGCTGCTGAGGAGCGGATATATGCGGAGCTC 15411
3541 CTGATGAGAGCGGCTTCTGCAACATCGGCGTCTGTCGCTGAGGAGCAATTCATTTT 3600
15412 CTGATGAGAGCGGCTTCTGCAACATCGGCGTCTGTCGCTGAGGAGCAATTCATTTT 15471
3601 GAAAGGTTGCGCGGCTTCTGAGCTGAGCAATTCGAGCTTTTACGTCAGGAGTGTG 3660
15472 GAAAGGTTGCGCGGCTTCTGAGCTGAGCAATTCGAGCTTTTACGTCAGGAGTGTG 15531
3661 GCGCGGAGGATGAGCGGCGGAGTTCAGGATGATGAGCTGCGGAGAGATTCCTCACCG 3720
15532 GCGCGGAGGATGAGCGGCGGAGTTCAGGATGATGAGCTGCGGAGAGATTCCTCACCG 15591
3721 AGCGCGGAGAGAGCGCGCGGCTTCGCGCGGAGAGCACTTCGCGATATGCTT 3780


```
Db 13072 GCGCGGCTCTGGGGATCCAGAGCGCATGTGTCCTGGTGCTCAGAGCGCGCTT 13131
Qy 1261 AACGAGCAAGTCGTTGGTGTCACCTCGTTGAGAGGCTCGGAACTCCGGTGTACACAGC 1320
Db 13132 AACGAGCAAGTCGTTGGTGTCACCTCGTTGAGAGGCTCGGAACTCCGGTGTACACAGC 13191
Qy 1321 ACGGAGACTCCTCAGCTGCTGCTGATCATCAGCTCTACGAGCAGATGGGAGCTCGTC 1380
Db 13192 ACGGAGACTCCTCAGCTGCTGCTGATCATCAGCTCTACGAGCAGATGGGAGCTCGTC 13251
Qy 1381 CTGCGTGGGACATCGTCGAGCGAGTGTTCGCGCCGACCTTCTGAGCAGACGTCTGAA 1440
Db 13252 CTGCGTGGGACATCGTCGAGCGAGTGTTCGCGCCGACCTTCTGAGCAGACGTCTGAA 13311
Qy 1441 GCGTACGTCGTTTCTCGCGGCGCTCATAGAGAACATGGGGTGAACAGGTGCGCTGT 1500
Db 13312 GCGTACGTCGTTTCTCGCGGCGCTCATAGAGAACATGGGGTGAACAGGTGCGCTGT 13371
Qy 1501 TCGCTTCGCTGCGCCAGCTAGAAAGCGGGCGAGCGGAAAGCGGACGAGCGCTGCTG 1560
Db 13372 TCGCTTCGCTGCGCCAGCTAGAAAGCGGGCGAGCGGAAAGCGGACGAGCGCTGCTG 13431
Qy 1561 AGCGAGCATACGCTGCAAGCGCTGTTCGCGCGCGGTCGAGCAGCTGCCATGACGCTC 1620
Db 13432 AGCGAGCATACGCTGCAAGCGCTGTTCGCGCGCGGTCGAGCAGCTGCCATGACGCTC 13491
Qy 1621 GCGCGTGTGTGCGCGCGGCAAGACGCTCATGTCAGAAAGCTTTGCGCGCTTCGCGGGA 1680
Db 13492 GCGCGTGTGTGCGCGCGGCAAGACGCTCATGTCAGAAAGCTTTGCGCGCTTCGCGGGA 13551
Qy 1681 CTTGGCGCGCGCTGCGCGAGCAGGGGAGCGCCGGAACATTTGTCGCGGTGTGATG 1740
Db 13552 CTTGGCGCGCGCTGCGCGAGCAGGGGAGCGCCGGAACATTTGTCGCGGTGTGATG 13611
Qy 1741 GAGAAAGCTGGAGAGCAGGTTGTGCGGTTCTGCGGTGCTGAGTCAGGCGCGCTTAC 1800
Db 13612 GAGAAAGCTGGAGAGCAGGTTGTGCGGTTCTGCGGTGCTGAGTCAGGCGCGCTTAC 13671
Qy 1801 GTGCGCATTCGATGCGGACCTTACCGCGGAGAGGCTATCCATCTCTCTGATCATGTGTAG 1860
Db 13672 GTGCGCATTCGATGCGGACCTTACCGCGGAGAGGCTATCCATCTCTCTGATCATGTGTAG 13731
Qy 1861 GTAAAGCTGTCGTGTCAGCAGCAGCATGCTGATGAGCAACTGTCAATGCGCGCGGATC 1920
Db 13732 GTAAAGCTGTCGTGTCAGCAGCAGCATGCTGATGAGCAACTGTCAATGCGCGCGGATC 13791
Qy 1921 CAGCGGCTGCTGTGAGCGAGGCGCGGCTGCAAGGCGACGCGCAGCTCCGATGATG 1980
Db 13792 CAGCGGCTGCTGTGAGCGAGGCGCGGCTGCAAGGCGACGCGCAGCTCCGATGATG 13851
Qy 1981 CCCATTGAGACACCTTCGGAATCTCGCGTATGTCTATCACTCTGCGGATTCACAGGGTTG 2040
Db 13852 CCCATTGAGACACCTTCGGAATCTCGCGTATGTCTATCACTCTGCGGATTCACAGGGTTG 13911
Qy 2041 CCCAAGGGGGTGTATGATCATCATCGGGGTGCGGCAACCATCTGGAATTCACAGG 2100
Db 13912 CCCAAGGGGGTGTATGATCATCATCGGGGTGCGGCAACCATCTGGAATTCACAGG 13971
Qy 2101 CGCTTCGAATAGGCGCCGAGACAGGGTGTGCGCTCTCTCGCTGAGCTTCGATCTC 2160
Db 13972 CGCTTCGAATAGGCGCCGAGACAGGGTGTGCGCTCTCTCGCTGAGCTTCGATCTC 14031
Qy 2161 TCGGCTATGATGTGTGGGATCTTGCGCGCGCGGCTGATGATGTGTGCGGACGCG 2220
Db 14032 TCGGCTATGATGTGTGGGATCTTGCGCGCGCGGCTGATGATGTGTGCGGACGCG 14091
Qy 2221 TCGAAGCTGCGGATTCGGCGCATTTGGCAGAGTTGATGAAAGAGAGGTGACGCTG 2280
Db 14092 TCGAAGCTGCGGATTCGGCGCATTTGGCAGAGTTGATGAAAGAGAGGTGACGCTG 14151
Qy 2281 TGGAACTGCGGTGCGCGCTGATGTGCTGTGAGCAATTTGAGGCTGCGCCGAT 2340
Db 14152 TGGAACTGCGGTGCGCGCTGATGTGCTGTGAGCAATTTGAGGCTGCGCCGAT 14211
Qy 2341 TCGCTCGCTAGGTCCTGCGCGCTTTCGCTGTGAGCGGCACTGATCCCGTGGGCTG 2400
Db 14212 TCGCTCGCTAGGTCCTGCGCGCTTTCGCTGTGAGCGGCACTGATCCCGTGGGCTG 14271
Qy 2401 CTTGGCAGCTCCAGGCACTAGGCGCGCGTGTGCGTGTATCAGCTTGGCGGGGCAAC 2460
Db 14272 CTTGGCAGCTCCAGGCACTAGGCGCGCGTGTGCGTGTATCAGCTTGGCGGGGCAAC 14331
Qy 2461 GAAGCTGATCTGTGTCATGCGGATACCCGTGAGAAAGCTTCGACTATCTGTGGCGAGC 2520
Db 14332 GAAGCTGATCTGTGTCATGCGGATACCCGTGAGAAAGCTTCGACTATCTGTGGCGAGC 14391
Qy 2521 ATCCCTACCGGCGGTGCGCGGCAAGAGCTTCCACCGTCTGATGAGGCGCTGAA 2580
Db 14392 ATCCCTACCGGCGGTGCGCGGCAAGAGCTTCCACCGTCTGATGAGGCGCTGAA 14451
Qy 2581 CCGCGCGGCTGTGGGTTCCGGGCAACTTACATTTGGCGGGTTCGGGCTGGCACTGGGC 2640
Db 14452 CCGCGCGGCTGTGGGTTCCGGGCAACTTACATTTGGCGGGTTCGGGCTGGCACTGGGC 14511
Qy 2641 TACTGGCGCATGAAAGAGAGAGCGGCAAGACTTCTCTGTGACCCCGAGACCGGGGAG 2700
Db 14512 TACTGGCGCATGAAAGAGAGAGCGGCAAGACTTCTCTGTGACCCCGAGACCGGGGAG 14571
Qy 2701 CGCCTTACAAAGCGCGGATCTGGGCGGCTACCTGCGGATGGAACATGAGTTATG 2760
Db 14572 CGCCTTACAAAGCGCGGATCTGGGCGGCTACCTGCGGATGGAACATGAGTTATG 14631
Qy 2761 GGGCGTGAAGCAACCAATCAAGCTTGCAGATACCGCGTTGAGCTCGGGAATCGAG 2820
Db 14632 GGGCGTGAAGCAACCAATCAAGCTTGCAGATACCGCGTTGAGCTCGGGAATCGAG 14691
Qy 2821 GAACGCTCAAGTTCGATTCGGAACGTACGAGCGCGGTGATTTGCCGTGCGGAAGAC 2880
Db 14692 GAACGCTCAAGTTCGATTCGGAACGTACGAGCGCGGTGATTTGCCGTGCGGAAGAC 14751
Qy 2881 GCGCGGAAAGAGCTCTTACCTATGAGTGTCCGGAAGGCAACAGGAGCGGGCTGC 2940
Db 14752 GCGCGGAAAGAGCTCTTACCTATGAGTGTCCGGAAGGCAACAGGAGCGGGCTGC 14811
Qy 2941 GAGCAGAGCGGAGCTTCAAGACCGAGCGGATTCGACGCGAGGACAGCGCCGGAAGCG 3000
Db 14812 GAGCAGAGCGGAGCTTCAAGACCGAGCGGATTCGACGCGAGGACAGCGCCGGAAGCG 14871
Qy 3001 GACGCTTGAAGCGGCGAGAGGGTCAAGTTCAGCTCGTCAACAGCATTCGGAAG 3060
Db 14872 GACGCTTGAAGCGGCGAGAGGGTCAAGTTCAGCTCGTCAACAGCATTCGGAAG 14931
Qy 3061 GACCTGAGCGGAAAGCGCGTGTGATCTGACCGGGGAGGATTCGCGGGAGCGGGGCTG 3120
Db 14932 GACCTGAGCGGAAAGCGCGTGTGATCTGACCGGGGAGGATTCGCGGGAGCGGGGCTG 14991
Qy 3121 GACGCTACGCGCGTTCGCGCGTGTGAGCGTTCGGAAGCTTCTTGAAGCGCGGATTC 3180
Db 14992 GACGCTACGCGCGTTCGCGCGTGTGAGCGTTCGGAAGCTTCTTGAAGCGCGGATTC 15051
Qy 3181 GAGTTGATGATTCCTGAGCTGTTAGCAGCGTGAAGCCGAGCGGCGGACCTTCTCC 3240
Db 15052 GAGTTGATGATTCCTGAGCTGTTAGCAGCGTGAAGCCGAGCGGCGGACCTTCTCC 15111
Qy 3241 AAATTCGTTATTCATGCGGCGGAGAGCATACCGGTTGAAACCTTACGCTATGTCAA 3300
Db 15112 AAATTCGTTATTCATGCGGCGGAGAGCATACCGGTTGAAACCTTACGCTATGTCAA 15171
Qy 3301 TCGGCGCATTCGAGGCGTGTGAGAGGCTTCTATTAATTAACACCGGTTGAGACAGCT 3360
Db 15172 TCGGCGCATTCGAGGCGTGTGAGAGGCTTCTATTAATTAACACCGGTTGAGACAGCT 15231
Qy 3361 TTGCTGAAGCTTTCGATTCACGCGGATTCAGCGCGGAGCGGACGTTGCGCAAACTTGAC 3420
Db 15232 TTGCTGAAGCTTTCGATTCACGCGGATTCAGCGCGGAGCGGACGTTGCGCAAACTTGAC 15291
```



```

QY 3421 GTGTTGATGAAGCGCGTTTCAACTCTGTTGTCGGGAGATGCAAGCCATCGAGTCG 3480
DB 15292 GTGTTGATGAAGCGCGCGTTTCAACTCTGTTGTCGGGAGATGCAAGCCATCGAGTCG 15351
QY 3481 CTGTAATGATCGCGTCGGGAGAAATTTGCGTCGTCGAGAGCGCGGAAATATATGCGGAGCTC 3540
DB 15352 CTGTAATGATCGCGTCGGGAGAAATTTGCGTCGTCGAGAGCGCGGAAATATATGCGGAGCTC 15411
QY 3541 CTGATGAGACAGGCGCGCTTCTGTCACATCGCGTCGTCGTCGGTGGGCAATTCATTTT 3600
DB 15412 CTGATGAGACAGGCGCGCTTCTGTCACATCGCGTCGTCGTCGGTGGGCAATTCATTTT 15471
QY 3601 GAAACAGTTTCGCGCGGTTCTTCCGACCTGCGACATTCGACGTTTACGTCACCGCATGCTG 3660
DB 15472 GAAACAGTTTCGCGCGGTTCTTCCGACCTGCGACATTCGACGTTTACGTCACCGCATGCTG 15531
QY 3661 GCGCGGCGGAGTGAACCGCGGAGATTCAGAGTCTGTAAGCTCGTGAAGATTCTTCACCG 3720
DB 15532 GCGCGGCGGAGTGAACCGCGGAGATTCAGAGTCTGTAAGCTCGTGAAGATTCTTCACCG 15591
QY 3721 AGCGCGCCACGACGCGCGCGCGCCCTCCGCGCGGAGACGACATTCGCGATATGCTT 3780
DB 15592 AGCGCGCCACGACGCGCGCGCGCCCTCCGCGCGGAGACGACATTCGCGATATGCTT 15651
QY 3781 CGCGACTTCTTGAAGACCAATCAACCGAGTACATGTCCTTACAGTCTTCTGAGAGCTC 3840
DB 15652 CGCGACTTCTTGAAGACCAATCAACCGAGTACATGTCCTTACAGTCTTCTGAGAGCTC 15711
QY 3841 GATGCGTTCGCGTGAAGTCAACCGGCAAGGATGATGTAAGGCGCTGCGCGAGCGGAG 3900
DB 15712 GATGCGTTCGCGTGAAGTCAACCGGCAAGGATGATGTAAGGCGCTGCGCGAGCGGAG 15771
QY 3901 GATACCTCTGCGCGCGGCAATTCGGGCAACGCGCGCAACGCGACCTTGAAGAGATC 3960
DB 15772 GATACCTCTGCGCGCGGCAATTCGGGCAACGCGCGCAACGCGACCTTGAAGAGATC 15831
QY 3961 CTGCTGCGCGCTGTAAGCGGAGTCTCGGCTGAGAGTGTGCGGCTTCCAGCAGAGCTTC 4020
DB 15832 CTGCTGCGCGCTGTAAGCGGAGTGTGCGGCTGAGAGTGTGCGGCTTCCAGCAGAGCTTC 15891
QY 4021 GTGCGATCTTGTGTCGATCGATTCATCGTTCGATGAGAGCGCTGTCGAGAGAG 4080
DB 15892 GTGCGATCTTGTGTCGATCGATTCATCGTTCGATGAGAGCGCTGTCGAGAGAG 15951
QY 4081 CTGAGTAGGAGATCGCATCAACGAGTGTTCAGTACCCGAACTCGGCTCGCTGCGG 4140
DB 15952 CTGAGTAGGAGATCGCATCAACGAGTGTTCAGTACCCGAACTCGGCTCGCTGCGG 16011
QY 4141 TCCGCTTTCGCGCGGACATCTGAGAGATCTTGAATCAAGCGCGCGAACTGAGAGCCGAGTG 4200
DB 16012 TCCGCTTTCGCGCGGACATCTGAGAGATCTTGAATCAAGCGCGCGAACTGAGAGCCGAGTG 16071
QY 4201 GAGGTTCCGCGCAAGGCGAGAGAGCTAGCTAA 4233
DB 16072 GAGGTTCCGCGCAAGGCGAGAGAGCTAGCTAA 16104

```

```

RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; CURRENT FILING DATE: 2000-05-10

```

```

; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1

Query Match      100.0%; Score 4233; DB 4; Length 68750;
Beef Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGATCAATCAGCTTCTGAACGAGCTCGAGACGAGGATATCAAGCTGCGCGCAT 60
DB 11872 ATGAGATCAATCAGCTTCTGAACGAGCTCGAGACGAGGATATCAAGCTGCGCGCAT 11931
QY 61 GGGGAGCGGCTCCAGATACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
DB 11932 GGGGAGCGGCTCCAGATACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11991
QY 121 ATCTCGAGCACAAAAGACGATCTGACGATGCTCCGTCAGAGACTCCCGCAGAAATCC 180
DB 11992 ATCTCGAGCACAAAAGACGATCTGACGATGCTCCGTCAGAGACTCCCGCAGAAATCC 12051
QY 181 ATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
DB 12052 ATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12111
QY 241 TACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 12112 TACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12171
QY 301 TACGACTGAGGATCTGACGCTGCGAGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 12172 TACGACTGAGGATCTGACGCTGCGAGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 12231
QY 361 CGGACGACATGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
DB 12232 CGGACGACATGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12291
QY 421 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
DB 12292 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12351
QY 481 AGGCTCGTGTGTCGAGATGCGATGTCGACCGCATCTATGACACCGAGCGCGCTCGG 540
DB 12352 AGGCTCGTGTGTCGAGATGCGATGTCGACCGCATCTATGACACCGAGCGCGCTCGG 12411
QY 541 CTCTATCAAGTGTGTCGCGTTCGCTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
DB 12412 CTCTATCAAGTGTGTCGCGTTCGCTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 12471
QY 601 GATCTCAATTAAGGATGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 12472 GATCTCAATTAAGGATGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12531
QY 661 TACGAGATCCGAGACCTCTCTCCCTGTCTGAGAGCTCTGTAACCGGATTAATGACTC 720
DB 12532 TACGAGATCCGAGACCTCTCTCCCTGTCTGAGAGCTCTGTAACCGGATTAATGACTC 12591
QY 721 GCGCTGAGTCTCGGAGAGAGTCTGAGCGGATGCAAGATGATGATTAATCGAGAGGG 780
DB 12592 GCGCTGAGTCTCGGAGAGAGTCTGAGCGGATGCAAGATGATGATTAATCGAGAGGG 12651
QY 781 CGCATCGCGAGACTCCACCTCCGCGAGCGCTTCGATGAGAGCGGATCTCACTCCCTG 840
DB 12652 CGCATCGCGAGACTCCACCTCCGCGAGCGCTTCGATGAGAGCGGATCTCACTCCCTG 12711
QY 841 AAGGAGATCCGCTTCGCGGACACGAGGATGAGTGCCTGCGAGACTCTCGGCGTGCATG 900
DB 12712 AAGGAGATCCGCTTCGCGGACACGAGGATGAGTGCCTGCGAGACTCTCGGCGTGCATG 12771

```


QY 901 AAGCGGCTGTTCGGGAGCGGGGCTGAACCCGACGGGCGTATCTGCTGCAATTTTCC 960
Db 12772 AAGCGGCTGTTCGGGAGCGGGGCTGAACCCGACGGGCGTATCTGCTGCAATTTTCC 12831
QY 961 GAGGTGATCGGGGCGCTGGAGGCGGAGCCCGGGTTTACGCTCAACAATAACGCTTCAAC 1020
Db 12832 GAGGTGATCGGGGCGCTGGAGGCGGAGCCCGGGTTTACGCTCAACAATAACGCTTCAAC 12891
QY 1021 CGGCTCCCCGTCATCCGGGGTGAAAGATATCAACGGGGGACTTCAAGTCATGCTCTC 1080
Db 12882 CGGCTCCCCGTCATCCGGGGTGAAAGATATCAACGGGGGACTTCAAGTCATGCTCTC 12951
QY 1081 CTGACATCGACACCACTCGGACAAAGAGCTTCAAGACGCGCTCAAGCTTCAAGAG 1140
Db 12952 CTGACATCGACACCACTCGGACAAAGAGCTTCAAGACGCGCTCAAGCTTCAAGAG 13011
QY 1141 CAGCTGTGGGAAAGCATGATATCATCTGCAAGTAAAGGCTTCAAGCTTCAAGAGGCGC 1200
Db 13012 CAGCTGTGGGAAAGCATGATATCATCTGCAAGTAAAGGCTTCAAGCTTCAAGAGGCGC 13071
QY 1201 GCCCGGCTCTGGGGATCCAAAGAGGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTT 1260
Db 13072 GCCCGGCTCTGGGGATCCAAAGAGGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTT 13131
QY 1261 AACCGAAGTCTGTGTGTCACTGTTGCAAGAGCTCGGAACTCCGCTGTACACAGC 1320
Db 13132 AACCGAAGTCTGTGTGTCACTGTTGCAAGAGCTCGGAACTCCGCTGTACACAGC 13191
QY 1321 ACGGAGACTCTCAGCTGCTGTGATCATGACTTCAAGAGCATGATGGGAGCCTGCTC 1380
Db 13192 ACGGAGACTCTCAGCTGCTGTGATCATGACTTCAAGAGCATGATGGGAGCCTGCTC 13251
QY 1381 CTCGGGTGGGACATGTGACAGAGGTGTTCCCGCCGACCTTTCGAGAGCATGTGCA 1440
Db 13252 CTCGGGTGGGACATGTGACAGAGGTGTTCCCGCCGACCTTTCGAGAGCATGTGCA 13311
QY 1441 GCGTACGTCTGTTTTCTCCGGGCGCTCACTGAGAACATGAGGAGTGAACAGGTGCGCTG 1500
Db 13312 GCGTACGTCTGTTTTCTCCGGGCGCTCACTGAGAACATGAGGAGTGAACAGGTGCGCTG 13371
QY 1501 TCGCTTCCGCTCCCACTAGAACGCGGGGAGCGCAACGCGACCAAGCGCTGCTG 1560
Db 13372 TCGCTTCCGCTCCCACTAGAACGCGGGGAGCGCAACGCGACCAAGCGCTGCTG 13431
QY 1561 AGCGAGCATAGCTGCAAGCGCTGTTCCGGGCGGGTTCAGACGCTCCCATGAGCTC 1620
Db 13432 AGCGAGCATAGCTGCAAGCGCTGTTCCGGGCGGGTTCAGACGCTCCCATGAGCTC 13491
QY 1621 GCCGTGTGTGCGGCGCGCAAGCGCTCACTGTAAGAGCTTTGCGCGCTTCCGCGGCA 1680
Db 13492 GCCGTGTGTGCGGCGCGCAAGCGCTCACTGTAAGAGCTTTGCGCGCTTCCGCGGCA 13551
QY 1681 CTTGCGCGGCGCTGCGGAGCAGGGGCAAGCCCGAAACAATTGTCGCGTGTGATG 1740
Db 13552 CTTGCGCGGCGCTGCGGAGCAGGGGCAAGCCCGAAACAATTGTCGCGTGTGATG 13611
QY 1741 GAGAAAGGCTGGGAGCAGGTTGTGCGGCTTCTCGGCTGTGAGTCAAGGCGCGCTTAC 1800
Db 13612 GAGAAAGGCTGGGAGCAGGTTGTGCGGCTTCTCGGCTGTGAGTCAAGGCGCGCTTAC 13671
QY 1801 GTGCGATCGATGCGGACCTTACCGGCGGAGCTTACCTACCTCTGATCATGATGATG 1860
Db 13672 GTGCGATCGATGCGGACCTTACCGGCGGAGCTTACCTACCTCTGATCATGATGATG 13731
QY 1861 GTAAAGCTGTGTGTCGACGACCATGCTGATGCAACTGTCAATGCGCGCGGAGATC 1920
Db 13732 GTAAAGCTGTGTGTCGACGACCATGCTGATGCAACTGTCAATGCGCGCGGAGATC 13791
QY 1921 CAGGCGCTGTGTGAGGAGGCGCGCTGCAAGGCGACGCGACCACTTCCGATGATG 1980
Db 13792 CAGGCGCTGTGTGAGGAGGCGCGCTGCAAGGCGACGCGACCACTTCCGATGATG 13851

QY 1981 CCATTCGACACCTTCGGATCTGCGGTATGTCATTCACCTCGGATCCAGAGGTTG 2040
Db 13852 CCATTCGACACCTTCGGATCTGCGGTATGTCATTCACCTCGGATCCAGAGGTTG 13911
QY 2041 CCCAAGGGGATGATTCATTCGAGGTCCTTCAACATCTTGGATATCAAGAG 2100
Db 13912 CCCAAGGGGATGATTCATTCGAGGTCCTTCAACATCTTGGATATCAAGAG 13971
QY 2101 CGCTTCGAATATAGGCGCGGAGCAGGGTGTGCGCTCTCTGCTGAGCTTCAATTC 2160
Db 13972 CGCTTCGAATATAGGCGCGGAGCAGGGTGTGCGCTCTCTGCTGAGCTTCAATTC 14031
QY 2161 TCGGTCTATATGATGTTTCGGATCTGCGCGGCGGTACGATGTTGTCGCGAGCG 2220
Db 14032 TCGGTCTATATGATGTTTCGGATCTGCGCGGCGGTACGATGTTGTCGCGAGCG 14091
QY 2221 TCAGAGCTGCGGATTCGGGCGCATTTGGGCAAGTTGATCGAACGAGAGAGTGAAGG 2280
Db 14092 TCAGAGCTGCGGATTCGGGCGCATTTGGGCAAGTTGATCGAACGAGAGAGTGAAGG 14151
QY 2281 TGGAACTCGGTCGCGGCGGTGATGCGGATGCTCGTCAAGCATTTTGAAGGTCGCGCGAT 2340
Db 14152 TGGAACTCGGTCGCGGCGGTGATGCGGATGCTCGTCAAGCATTTTGAAGGTCGCGCGAT 14211
QY 2341 TCGCTCGTATAGTCTTTCGGCTTTCGCTGTCGAGCGGAGCTGATTCGCGTGGCCTG 2400
Db 14212 TCGCTCGTATAGTCTTTCGGCTTTCGCTGTCGAGCGGAGCTGATTCGCGTGGCCTG 14271
QY 2401 CTTGCGAGCTCGAGGCGCATGAGCCCGGCGGTTCGCTGATCAGCTTGGGCGGCGCAC 2460
Db 14272 CTTGCGAGCTCGAGGCGCATGAGCCCGGCGGTTCGCTGATCAGCTTGGGCGGCGCAC 14331
QY 2461 GAAAGCTGATCTGCTCATGCGGATACCCGTTGAGGAAAGCTGACCTATCGTGGGCGAGC 2520
Db 14332 GAAAGCTGATCTGCTCATGCGGATACCCGTTGAGGAAAGCTGACCTATCGTGGGCGAGC 14391
QY 2521 ATCCCTTACCGGCGCTGCTGCGGCAACGAGCTTTCACGCTGATGAGGCGCTGCA 2580
Db 14392 ATCCCTTACCGGCGCTGCTGCGGCAACGAGCTTTCACGCTGATGAGGCGCTGCA 14451
QY 2581 CCGGCGCGGCTGTGGGTTCCGGGCAACTTACATTTGGCGGGTTCGGCTGCGCTG 2640
Db 14452 CCGGCGCGGCTGTGGGTTCCGGGCAACTTACATTTGGCGGGTTCGGCTGCGCTG 14511
QY 2641 TACTGGGCGATGGAAGAGAGCGGCAAGAGCTTCCCTCGTGCACCCCGAGACCGGGAG 2700
Db 14512 TACTGGGCGATGGAAGAGAGCGGCAAGAGCTTCCCTCGTGCACCCCGAGACCGGGAG 14571
QY 2701 CGCCTTACAAAGCCGGCGATCTGGGCGCTTACCTGCGGATGGAACATCGAGTTCA 2760
Db 14572 CGCCTTACAAAGCCGGCGATCTGGGCGCTTACCTGCGGATGGAACATCGAGTTCA 14631
QY 2761 GGGGCTGAGCAACAATCAAGCTTTCGCGATATCCGCTTGAAGCTTGGGAAATCGAG 2820
Db 14632 GGGGCTGAGCAACAATCAAGCTTTCGCGATATCCGCTTGAAGCTTGGGAAATCGAG 14691
QY 2821 GAAACGCTCAAGTGCATTCGGAAGCTTACGCAAGCGGCTGATTTGTCGCTGGGAAAGC 2880
Db 14692 GAAACGCTCAAGTGCATTCGGAAGCTTACGCAAGCGGCTGATTTGTCGCTGGGAAAGC 14751
QY 2881 GCGGCGAAAGCTCTTCTTATGCTTATGTTGCTCCGAGAGGCAACAGGAGCGCTG 2940
Db 14752 GCGGCGAAAGCTCTTCTTATGCTTATGTTGCTCCGAGAGGCAACAGGAGCGCTG 14811
QY 2941 GAGCAGGACGAGCTTCAAGACGAGCGGATTCGACGCGAGAGCAACGCGCGAGACG 3000
Db 14812 GAGCAGGACGAGCTTCAAGACGAGCGGATTCGACGCGAGAGCAACGCGCGAGACG 14871
QY 3001 GACGCGTTGAGCGACGCGGAGGCTGCAAGTCAAGCTTCCGCAACGAGCTTCCGAG 3060
Db 14872 GACGCGTTGAGCGACGCGGAGGCTGCAAGTCAAGCTTCCGCAACGAGCTTCCGAG 14931
QY 3061 GACCTGAGCGGAAAGCCGCTGTGATCTGACCGGAGAGATCCGCGGAGCGGCGCTG 3120

Db	12412	CTCTATCAACGTCGCGCGCTTCGGCTGGAAGAGGGCAAAACCCGTCCTGGCTCAAGTATC	12471
OY	601	GATTCATTAAAGTTGACCTAGGAGAGCTCTGCATCATCTTCAAGAGACTGGCTCAAGCTTC	660
Db	12472	GATTCATTAAAGTTGACCTAGGAGAGCTCTGCATCATCTTCAAGAGACTGGCTCAAGCTTC	12531
OY	661	TACGAGATCCCGAGACCTCTCTCTCCCTGTCCTGAGAGCTCTCGTACCGCATTAATGATCTC	720
Db	12532	TACGAGATCCCGAGACCTCTCTCTCCCTGTCCTGAGAGCTCTCGTACCGCATTAATGATCTC	12591
OY	721	GCGCTGGAAGTCTCGCAAGAAGTCTGAGAGCGCATCAACGATCGATGGAATTACTGGAAGCGG	780
Db	12592	GCGCTGGAAGTCTCGCAAGAAGTCTGAGAGCGCATCAACGATCGATGGAATTACTGGAAGCGG	12651
OY	781	CGCATCCCGGAGCTCCCAACCTCCGCGCAGACGCTTCCGATGGAAGCCGATCCATCACTCCTG	840
Db	12652	CGCATCCCGGAGCTCCCAACCTCCGCGCAGACGCTTCCGATGGAAGCCGATCCATCACTCCTG	12711
OY	841	AAGGAGATCCGCTTCCGGCAACGAGACAAATGACTGCGCTCGGAGCTCTGGGGTGCATTTG	900
Db	12712	AAGGAGATCCGCTTCCGGCAACGAGACAAATGACTGCGCTCGGAGCTCTGGGGTGCATTTG	12771
OY	901	AAGCGGCGTGTCCGGGGAGCGCGGGCTTGAACCCCGACCGGGCGTCACTCTGGCTGCATTTTCC	960
Db	12772	AAGCGGCGTGTCCGGGGAGCGCGGGCTTGAACCCCGACCGGGCGTCACTCTGGCTGCATTTTCC	12831
OY	961	GAGGTGATCCGGGAGCTCTGAGAGCGGAGCGCCCGGTTTACGCTCAACAATACGCTCTTCAAC	1020
Db	12832	GAGGTGATCCGGGAGCTCTGAGAGCGGAGCGCCCGGTTTACGCTCAACAATACGCTCTTCAAC	12891
OY	1021	CGGCTCCCGCTCCATCCGCGCGGTGAACGATATCACCGGGAGCTTCAACGTCGATGATCTCTC	1080
Db	12892	CGGCTCCCGCTCCATCCGCGCGGTGAACGATATCACCGGGAGCTTCAACGTCGATGATCTCTC	12951
OY	1081	CTGGAACATCCGACACCACTCCGCGACAAAGACTTGAACAGCGCGCTTAAGGTATTCAAGG	1140
Db	12952	CTGGAACATCCGACACCACTCCGCGACAAAGACTTGAACAGCGCGCTTAAGGTATTCAAGG	13011
OY	1141	CAGCTGTGGGAAGGAGATGATATCACTGGAAGTAAAGGGATTCAGAGGTCCAGCGAGAGGAGC	1200
Db	13012	CAGCTGTGGGAAGGAGATGATATCACTGGAAGTAAAGGGATTCAGAGGTCCAGCGAGAGGAGC	13071
OY	1201	GCCCGGGTCTCTGGGGATTCACAAGAGGCGCAATTGTTCCCGCTGGTGTCAAGAGCGGCTT	1260
Db	13072	GCCCGGGTCTCTGGGGATTCACAAGAGGCGCAATTGTTCCCGCTGGTGTCAAGAGCGGCTT	13131
OY	1261	AACCAAGCAAGTCGTTGGTGTCACTCGTTGCAAGAGGCTCGGAACCTCCGCTGTACACACAC	1320
Db	13132	AACCAAGCAAGTCGTTGGTGTCACTCGTTGCAAGAGGCTCGGAACCTCCGCTGTACACACAC	13191
OY	1321	ACGCAAGCTCCTCAAGCTGCTGCTGGAATCACTACTCTACAGAGACGATGGGACCTCCGTC	1380
Db	13192	ACGCAAGCTCCTCAAGCTGCTGCTGGAATCACTACTCTACAGAGACGATGGGACCTCCGTC	13251
OY	1381	CTCGCTGTGGGACATCGTGTGACGAGAGTGTCCCGCCGACCTTCTGGAAGCAATGCTCGAA	1440
Db	13252	CTCGCTGTGGGACATCGTGTGACGAGAGTGTCCCGCCGACCTTCTGGAAGCAATGCTCGAA	13311
OY	1441	GCGTACGTGTTTTTCTCCGGCGGCTCATCTGAGGAACCATGGGGTGAACAAGTGCCTGT	1500
Db	13312	GCGTACGTGTTTTTCTCCGGCGGCTCATCTGAGGAACCATGGGGTGAACAAGTGCCTGT	13371
OY	1501	TTCGTTCCCGCTGCGCCAGCTAGAAAGCGCGGGCGAGCGCAAAACCGCAACAAACCGCTGCTG	1560
Db	13372	TTCGTTCCCGCTGCGCCAGCTAGAAAGCGCGGGCGAGCGCAAAACCGCAACAAACCGCTGCTG	13431
OY	1561	AGCAGAGATACGCTGACAGGCTCTTGGCGGCGCGGGTGCAGAGAGCTGCCATGACAGCTC	1620
Db	13432	AGCAGAGATACGCTGACAGGCTCTTGGCGGCGCGGGTGCAGAGAGCTGCCATGACAGCTC	13491
OY	1621	GCCGTGCTGTCCGCGCAAGACGCTCAGCAAGAACGTTTTCCGCGCTTCCGCGCGA	1680

Db	13492	GCCGTGGTGTTCGGCGCCGCAAGACGCTCAAGTACGAAGAGCTTTCCGCGCCGTTCCGGCCGA	13553
QY	1681	CTTGGCGCGCGCTCGCGCGAGCAGGGGGCACGCCCGCAACATTTGGTCGCGGTGTGATG	1740
Db	13552	CTTGGCGCGCGGCTCGCGCGAGCAGGGGGCACGCCCGCAACATTTGGTCGCGGTGTGATG	13613
QY	1741	GAGAAAGGCTGGAGCAGGTTGTTCGGGTTCTCGCGTGTCTGATGACGCGCGCTTC	1800
Db	13612	GAGAAAGGCTGGAGCAGGTTGTTCGGGTTCTCGCGTGTCTGATGACGCGCGCTTC	13671
QY	1801	GTGGCGATGATGCGGCACTACCGGGGGAGCGTATCATCTACCTCTCGATCATATGGTGAAG	1860
Db	13672	GTGGCGATGATGCGGCACTACCGGGGGAGCGTATCATCTACCTCTCGATCATATGGTGAAG	13733
QY	1861	GTAAGGCTCGTGTGACGCGACCATGGCTGGATGGGAAACTGTGATAGCGCGCGGGATC	1920
Db	13732	GTAAGGCTCGTGTGACGCGACCATGGCTGGATGGGAAACTGTGATAGCGCGCGGGATC	13791
QY	1921	CAGCGGCTGCTCGTGAACGAGGCCGGGCTGGAAGCGGAGCGGCGACCAAGCTTCGATGATG	1980
Db	13792	CAGCGGCTGCTCGTGAACGAGGCCGGGCTGGAAGCGGAGCGGCGACCAAGCTTCGATGATG	13851
QY	1981	CCCAATTACAGACACCTTCGGATCTCGGGTATGATCATACACCTCCGGGAATCCACAGGGTTG	2040
Db	13852	CCCAATTACAGACACCTTCGGATCTCGGGTATGATCATACACCTCCGGGAATCCACAGGGTTG	13911
QY	2041	CCCAAGGGGGGTGATGATCATCATCGGGGGTCCGTCACACATCCTCGACATCAACGAG	2100
Db	13912	CCCAAGGGGGGTGATGATCATCATCGGGGGTCCGTCACACATCCTCGACATCAACGAG	13971
QY	2101	CGCTTCAAAATAGGGCCCGGAGACAGGGTGTGGCGCTCTCTCGCTGAGCTTCGATCTC	2160
Db	13972	CGCTTCAAAATAGGGCCCGGAGACAGGGTGTGGCGCTCTCTCTCGCTGAGCTTCGATCTC	14031
QY	2161	TCCGCTCATATATGTGTTCCGGGAATCCCTGCGCGCGGGCGGTATCCATCTCGTGGCCGAGCGG	2220
Db	14032	TCCGCTCATATATGTGTTCCGGGAATCCCTGCGCGCGGGCGGTATCCATCTCGTGGCGGAG	14091
QY	2221	TCCAAGCTGGCGCATCCGGCGCATTTGGGCGAGATTGATCGAACGAGAGAAGGTGACGTG	2280
Db	14092	TCCAAGCTGGCGCATCCGGCGCATTTGGGCGAGATTGATCGAACGAGAGAAGGTGACGTG	14151
QY	2281	TGGAACTCGGGTGCCGGCGCTGATGCGGATGCTGTCGAGCAATTTGAGGGTCCGCCGAT	2340
Db	14152	TGGAACTCGGGTGCCGGCGCTGATGCGGATGCTGTCGAGCAATTTGAGGGTCCGCCGAT	14211
QY	2341	TCCGCTCCCTAGGTCTCTCGCGGCTTTTGGCTGTCTGAGCGCGGACTGATATCCCGGTGGGCTTG	2400
Db	14212	TCCGCTCCCTAGGTCTCTCGCGGCTTTTGGCTGTCTGAGCGCGGACTGATATCCCGGTGGGCTTG	14271
QY	2401	CTTGGCGAGGCTCCAGGCGCATCAGGCGCGGGGTGTCGGTATCATGCTTGGGGCGGGGCGAC	2460
Db	14272	CTTGGCGAGGCTCCAGGCGCATCAGGCGCGGGGTGTCGGTATCATGCTTGGGGCGGGGCGAC	14331
QY	2461	GAAAGCTGATCTGTGTTCATCGGGTATCCCGTGAAGAGAGTGTGACCTTATCTGTGGCGAGC	2520
Db	14332	GAAAGCTGATCTGTGTTCATCGGGTATCCCGTGAAGAGAGTGTGACCTTATCTGTGGCGAGC	14391
QY	2521	ATCCCTTACGCGCGGTCCGCTGCGCAACAGAGCTTCTCATGCTGCTGATGAGCGCTCGAA	2580
Db	14392	ATCCCTTACGCGCGGTCCGCTGCGCAACAGAGCTTCTCATGCTGCTGATGAGCGCTCGAA	14451
QY	2581	CCGGCGCCGGGTCTGGGTTCCGGGGCAACTTACATTGGCGGGGTGGGCTGGCACTGGAG	2640
Db	14452	CCGGCGCCGGGTCTGGGTTCCGGGGCAACTTACATTGGCGGGGTGGGCTGGCACTGGAG	14511
QY	2641	TACTGCGCGATGAGAAGAGAGCGCGCAAGACTTCTGTGTGACCCCGAGACCGGGGAG	2700
Db	14512	TACTGCGCGATGAGAAGAGAGCGCGCAAGACTTCTGTGTGACCCCGAGACCGGGGAG	14571
QY	2701	CGCCTTACAAAGACCGGCGATCTTGGGCGGTACCTGCGGATCGGATGGAACATCGAGTTCATG	2760
Db	14572	CGCCTTACAAAGACCGGCGATCTTGGGCGGTACCTGCGGATCGGATGGAACATCGAGTTCATG	14631

QY 2761 GGGCGTGAAGACCAACCAATCAAGCTTCGCGGATACCAGCTTGAGCTCGGGGAAATCGAG 2820
DB 14632 GGGCGTGAAGACCAACCAATCAAGCTTCGCGGATACCAGCTTGAGCTCGGGGAAATCGAG 14691
QY 2821 GAAACCTCAAGTCGATCGAAGCTACGCAAGCGGGTGAATTGCGCCGTCGGGAACGAC 2880
DB 14692 GAAACCTCAAGTCGATCGAAGCTACGCAAGCGGGTGAATTGCGCCGTCGGGAACGAC 14751
QY 2881 GCGGCAACAAAGCTCTTCTAGCTATGATGTCCTCGAGGGGACACGAGACGCGCTGCC 2940
DB 14752 GCGGCAACAAAGCTCTTCTAGCTATGATGTCCTCGAGGGGACACGAGACGCGCTGCC 14811
QY 2941 GAGCAGAGACCGGAGCTTCAAGACCGGAGCGGATGAGACCGGAGACCAACCGCCGAAAGG 3000
DB 14812 GAGCAGAGACCGGAGCTTCAAGACCGGAGCGGATGAGACCGGAGACCAACCGCCGAAAGG 14871
QY 3001 GACGCGTTGAGCGACGCGGAGAGGGTGAAGTTCAAGCTCGCTGACACGGAATCCGGAAG 3060
DB 14872 GACGCGTTGAGCGACGCGGAGAGGGTGAAGTTCAAGCTCGCTGACACGGAATCCGGAAG 14931
QY 3061 GACCTGAGCGAAAGCCCGTCTGATCTGAACCGGACGAGATCCGCGGAGCGGGGCTG 3120
DB 14932 GACCTGAGCGAAAGCCCGTCTGATCTGAACCGGACGAGATCCGCGGAGCGGGGCTG 14991
QY 3121 GACGCTTACCGCGGTCGCGGAGCGGTCGAAAGCTTCCTGAGGCGCCGATTCGCTTTGTT 3180
DB 14992 GACGCTTACCGCGGTCGCGGAGCGGTCGAAAGCTTCCTGAGGCGCCGATTCGCTTTGTT 15051
QY 3181 GAGTTTGTGATTCCTGAGCTCTTGAAGCTTGAAGCGGTGAAGCCGCGGACGCTTCCC 3240
DB 15052 GAGTTTGTGATTCCTGAGCTCTTGAAGCGGTGAAGCCGCGGACGCTTCCC 15111
QY 3241 AAATTCGTTATTCATTCGCGGGGACACGTAACCGGTGAAACCTACCGGTATGTCAA 3300
DB 15112 AAATTCGTTATTCATTCGCGGGGACACGTAACCGGTGAAACCTACCGGTATGTCAA 15171
QY 3301 TCCGCGGATCGAAGGCGGTCGACGAGGCTTATATTAACAACCGCTTCGAGCAACCT 3360
DB 15172 TCCGCGGATCGAAGGCGGTCGACGAGGCTTATATTAACAACCGCTTCGAGCAACCT 15231
QY 3361 TTGCTGAAGCTCTCCGATCAAGGATCGAGCGGACGCACTTCGCGAAACTTCGAC 3420
DB 15232 TTGCTGAAGCTCTCCGATCAAGGATCGAGCGGACGCACTTCGCGAAACTTCGAC 15291
QY 3421 GTGTTGATGAGAGCGGCTTCAACTCTGTTCTGTTGGGAGATCGACGCTATCGAGTGC 3480
DB 15292 GTGTTGATGAGAGCGGCTTCAACTCTGTTCTGTTGGGAGATCGACGCTATCGAGTGC 15351
QY 3481 CTGATGAGATCGATCGGAGAAATTTGCTGCTGAGAGGCGGATATATGCGGAGCTC 3540
DB 15352 CTGATGAGATCGATCGGAGAAATTTGCTGCTGAGAGGCGGATATATGCGGAGCTC 15411
QY 3541 CTGATGAGAGCGGCTTCTGCAACATCGAGCTGTCTCGGTGGGAGCAATTCATTTT 3600
DB 15412 CTGATGAGAGCGGCTTCTGCAACATCGAGCTGTCTCGGTGGGAGCAATTCATTTT 15471
QY 3601 GAAACAGTTTCGCGCGGTTCTGCACTGCGACATTCGAGCTTTTACGTCAGCGAGCTGC 3660
DB 15472 GAAACAGTTTCGCGCGGTTCTGCACTGCGACATTCGAGCTTTTACGTCAGCGAGCTGC 15531
QY 3661 GCGGCGGCGGTAGACCGCGGAGTTTCAGGATGATGATGATGATGATGATGATGATGATG 3720
DB 15532 GCGGCGGCGGTAGACCGCGGAGTTTCAGGATGATGATGATGATGATGATGATGATGATG 15591
QY 3721 AGGCGCGGACGACGCGGCGGCGCTTCGCGCGGAGAGAGACTTCGCGGATGATGATG 3780
DB 15592 AGGCGCGGACGACGCGGCGGCGCTTCGCGCGGAGAGAGACTTCGCGGATGATGATG 15651
QY 3781 CGGAGCTTCTTGAAGACCAACTACCGGATGATGATGATGATGATGATGATGATGATGATG 3840
DB 15652 CGGAGCTTCTTGAAGACCAACTACCGGATGATGATGATGATGATGATGATGATGATGATG 15711

QY 3841 GATGCTTCCGCTGACGCTCAAGCGGCAAGGTCGATCGTAAAGGCCCTGCGGAGCGGAG 3900
DB 15712 GATGCTTCCGCTGACGCTCAAGCGGCAAGGTCGATCGTAAAGGCCCTGCGGAGCGGAG 15771
QY 3901 GATACCTGCTCGCGCGGATTCGCGGAGCAACGCGGCAACGCGGACCTTGGAGAGATC 3960
DB 15772 GATACCTGCTCGCGCGGATTCGCGGAGCAACGCGGCAACGCGGACCTTGGAGAGATC 15831
QY 3961 CTGCTGCGGCTGTACGCGGAGGTCCTCGGCTGAGAGGTGTGTGGCTTCAAGAGCTTC 4020
DB 15832 CTGCTGCGGCTGTACGCGGAGGTCCTCGGCTGAGAGGTGTGTGGCTTCAAGAGCTTC 15891
QY 4021 GTGATCTTGTGCGACATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 4080
DB 15892 GTGATCTTGTGCGACATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 15951
QY 4081 CTGATGAGGAGATCGCATCAACGAGTGTTCAGTACCCGAACTCGCTCGTGGCG 4140
DB 15952 CTGATGAGGAGATCGCATCAACGAGTGTTCAGTACCCGAACTCGCTCGTGGCG 16011
QY 4141 TCCGTTTCCGCGGAGATTCGAGAGATCTAAGTACGCGGCGGAAATGAGAGACGAGT 4200
DB 16012 TCCGTTTCCGCGGAGATTCGAGAGATCTAAGTACGCGGCGGAAATGAGAGACGAGT 16071
QY 4201 GAGTTCCGCGGAGGAGGAGAGAGTACGTA 4233
DB 16072 GAGTTCCGCGGAGGAGGAGAGAGTACGTA 16104

RESULT 8
US-09-443-501A-2

Sequence 2, Application US/09443501A
Patent No. 630342
GENERAL INFORMATION:
APPLICANT: Kosan Biosciences, Inc.
APPLICANT: Julien, Bryan
APPLICANT: Katz, Leonard
APPLICANT: Khosla, Chaitan
APPLICANT: Tang, Li
APPLICANT: Ziermann, Rainer
TITLE OF INVENTION: Recombinant Methods and Materials for Producing
TITLE OF INVENTION: Epochlone and Epochlone Derivatives
FILE REFERENCE: 30062-20031.00
CURRENT APPLICATION NUMBER: US/09/443,501A
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: US 60/130,560
PRIOR FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: US 60/122,620
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/119,386
PRIOR FILING DATE: 1998-02-10
PRIOR APPLICATION NUMBER: US 60/109,401
PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 71989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 97.2%; Score 4114.6; DB 4; Length 71989;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 4159; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGAGCATTAATCAGCTTCTGAACGAGCTCGAGCAACGAGGATTAAGCTGGCGGCGAT 60
DB 6260 ATGAGCATTAATCAGCTTCTGAACGAGCTCGAGCAACGAGGATTAAGCTGGCGGCGAT 6319
QY 61 GGGAGCGGCTCAATACAGGCGCCCAAGAGCGGCTGAACCCGAACTGCTGCTCGA 120

Dh 6320 GGGGAGCCCTTCAGATACAGGCCCCCAAGAACGCCCTGAAACCCGAACTGCTCGCTGA 6379
Qy 121 ATCTCCGAGCACAAAAGCAGCATCTTGAAGATGCTCCGTGAGAGATCTCCCGCAGAAATCC 180
Db 6380 ATCTCCGAGCACAAAAGCAGCATCTTGAAGATGCTCCGTGAGAGATCTCCCGCAGAAATCC 6439
Qy 181 ATCTGCCCCCGCCAGCCGAGGAGCAGCTCGTTCTCTTCAAGACATCCAAAGATCC 240
Db 6440 ATCTGCCCCCGCCAGCCGAGGAGCAGCTCGTTCTCTTCAAGACATCCAAAGATCC 6499
Qy 241 TACTGCGCGGCGCCGAGAGAGCTTTTACGTTCCCAAGGAGATCCAGCCCTATGCGGA 300
Db 6500 TACTGCGCGGCGCCGAGAGAGCTTTTACGTTCCCAAGGAGATCCAGCCCTATGCGGA 6559
Qy 301 TACGACTGTACGAGATCTGACGCTGCGAGAGCTGAGCCGCTTTCCGAAAGTCTGCGC 360
Db 6560 TACGACTGTACGAGATCTGACGCTGCGAGAGCTGAGCCGCTTTCCGAAAGTCTGCGC 6619
Qy 361 CGGACGACATGCTTCGCGGCCCAACGCTGCCGACATGATGCAAGGTATGAGCCTTAA 420
Db 6620 CGGACGACATGCTTCGCGGCCCAACGCTGCCGACATGATGCAAGGTATGAGCCTTAA 6679
Qy 421 GTCCAGCGCCGACATGAGATCATCTGCGCGGCGCTGACCGGAGACACAGGAGACG 480
Db 6680 GTCCAGCGCCGACATGAGATCATCTGCGCGGCGCTGACCGGAGACACAGGAGACG 6739
Qy 481 AGGCTGCTGCTTGCAGATGCGAGATGCGACCGCATCTATGACAGGAGCGCCCTCG 540
Db 6740 AGGCTGCTGCTTGCAGATGCGAGATGCGACCGCATCTATGACAGGAGCGCCCTCG 6799
Qy 541 CTCTATCACTGCTGCTGCGCTGCGCTGAGACAGCGGCAAAACCCGTCTGCTGCTAGTATC 600
Db 6800 CTCTATCACTGCTGCTGCGCTGCGCTGAGACAGGCAAAACCCGTCTGCTGCTAGTATC 6859
Qy 601 GATCTCATTTAAGCTTGAAGCTTGAAGAGCTTGTCTCATCTTCAAGAGATGAGTCACTTC 660
Db 6860 GATCTCATTTAAGCTTGAAGAGCTTGTCTCATCTTCAAGAGATGAGTCACTTC 6919
Qy 661 TACGAGATCCCGAGACCTCTCTCCGTGCTGAGCTCTGCTGAGGAGTTATGATCTC 720
Db 6920 TACGAGATCCCGAGACCTCTCTCCGTGCTGAGCTCTGCTGAGGAGTTATGATCTC 6979
Qy 721 GCGCTGAGCTGCGCAAGAGCTGAGGCGCATCAAGCATGAGTATTAATGAGAGCG 780
Db 6980 GCGCTGAGCTGCGCAAGAGCTGAGGCGCATCAAGCATGAGTATTAATGAGAGCG 7039
Qy 781 CGCATCCGCGAGCTCCACCTCCGCGACGCTTCCGATGAAGCGCATCTACCTG 840
Db 7040 CGCATCCGCGAGCTCCACCTCCGCGAGCTTCCGATGAAGCGCATCTACCTG 7099
Qy 841 AAGGAGATCCGCTTCCGCGACACGAGAGATGCGCTGCGGATCTCTGGGCTGATG 900
Db 7100 AAGGAGATCCGCTTCCGCGACACGAGAGATGCGCTGCGGATCTCTGGGCTGATG 7159
Qy 901 AAGCGCGCTGCTCGGAGAGCGCGGCTGACCCCGAGCGGCTGATCTGCTGATTTTC 960
Db 7160 AAGCGCGCTGCTCGGAGAGCGCGGCTGACCCCGAGCGGCTGATCTGCTGATTTTC 7219
Qy 961 GAGGTGATCGGCGCTGAGGCGCGAGCCCGGTTTACGCTTCAATTAACGCTTTCAAC 1020
Db 7220 GAGGTGATCGGCGCTGAGGCGCGAGCCCGGTTTACGCTTCAATTAACGCTTTCAAC 7279
Qy 1021 CGGCTCCCGCTGCTGCGGCTGAGAGATGACCGGAGGCTTCAAGTATGATGCTCTC 1080
Db 7280 CGGCTCCCGCTGCTGCGGCTGAGAGATGACCGGAGGCTTCAAGTATGATGCTCTC 7339
Qy 1081 CTGACATCGACACCACTCGCGACAAAGAGCTTCAAGAGCGGCTTCAAGAG 1140
Db 7340 CTGACATCGACACCACTCGCGACAAAGAGCTTCAAGAGCGGCTTCAAGAG 7399
Qy 1141 CAGCTGTGGAGAGCATGATGATCTGCGAGCTTGAAGCGGATATGAGTCCAGGAGAGCC 1200
Db 7400 CAGCTGTGGAGAGCATGATGATCTGCGAGCTTGAAGCGGATATGAGTCCAGGAGAGCC 7459

Qy 1201 GCCCGGCTCTGGGATTCACAGAGCGCATTTGTTCCCGGTGCTCAAGAGCGCGCTT 1260
Db 7460 GCCCGGCTCTGGGATTCACAGAGCGCATTTGTTCCCGGTGCTCAAGAGCGCGCTC 7519
Qy 1261 AACGAGAGCTGTTGCTGCTCACTCTGTTGAGAGGCTCGGAACTCCGTGTACACAGC 1320
Db 7520 AACGAGAGCTGTTGCTGCTCACTCTGTTGAGAGGCTCGGAACTCCGTGTACACAGC 7579
Qy 1321 ACGGAGCTCCTGAGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1380
Db 7580 ACGGAGCTCCTGAGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 7639
Qy 1381 CTCCGCTGGAGCATGCTGCAAGAGTGTCCCGCCGACCTTCTGAGACGATGCTCGAA 1440
Db 7640 CTCCGCTGGAGCATGCTGCAAGAGTGTCCCGCCGACCTTCTGAGACGATGCTCGAA 7699
Qy 1441 GCGTACGCTGTTTTTCTCCGCGGCTCACTGAGAAACATGGGAGTAAACAGTGCCTGT 1500
Db 7700 GCGTACGCTGTTTTTCTCCGCGGCTCACTGAGAAACATGGAGTAAACAGTGCCTGT 7759
Qy 1501 TCGCTTCCGCTGCGCAAGAGCGCGGCGAGCCGAAACGAGACCAACCTGCTGTG 1560
Db 7760 TCGCTTCCGCTGCGCAAGAGCGCGGCGAGCCGAAACGAGACCAACCTGCTGTG 7819
Qy 1561 AGCGAGCATGAGCTGCAAGGCTGTTGCGCGGCGGAGCTGAGAGCTGCCATGAGCTC 1620
Db 7820 AGCGAGCATGAGCTGCAAGGCTGTTGCGCGGCGGAGCTGAGAGCTGCCATGAGCTC 7879
Qy 1621 GCGGTGCTGCGGCGCAAGAGCGCTCACTGACAGAGCTTTGCGCGCTTTCGCGCGCA 1680
Db 7880 GCGGTGCTGCGGCGCAAGAGCGCTCACTGACAGAGCTTTGCGCGCTTTCGCGCGCA 7939
Qy 1681 CTTGGCGCGGCTGCGGCGAGCAGGCGGCAAGCCGAAACATTTGCTGCGGTGATG 1740
Db 7940 CTTGGCGCGGCTGCGGCGAGCAGGCGGCAAGCCGAAACATTTGCTGCGGTGATG 7999
Qy 1741 GAGAAAGCTGGAGAGAGTGTGCGGCTTCTCGCGGCTGAGTCAAGGCGCGGCTTAC 1800
Db 8000 GAGAAAGCTGGAGAGAGTGTGCGGCTTCTCGCGGCTGAGTCAAGGCGCGGCTTAC 8059
Qy 1801 GTGCGATGATGCGAGCTTACCGGCGAGAGCTTACCTGCTTCTGATCATGCTGAG 1860
Db 8060 GTGCGATGATGCGAGCTTACCGGCGAGAGCTTACCTGCTTCTGATCATGCTGAG 8119
Qy 1861 GTAAAGCTGCTGCTGAGAGCAGCAGCTGCTGATGAGCAACCTGCTGCTGCTGCTG 1920
Db 8120 GTAAAGCTGCTGCTGAGAGCAGCAGCTGCTGATGAGCAACCTGCTGCTGCTGCTG 8179
Qy 1921 CAGGCGCTGCTGCTGAGAGAGCGGCGGCTGAGAGCGGCGAGCAGGCTCGATGATG 1980
Db 8180 CAGGCGCTGCTGCTGAGAGAGCGGCGGCTGAGAGCGGCGAGCAGGCTCGATGATG 8239
Qy 1981 CCAATTCAGACACTTGGATCTTGGATGCTGATGCTTCACTTCCGAGATCCAGAGGTTG 2040
Db 8240 CCAATTCAGACACTTGGATCTTGGATGCTGATGCTTCACTTCCGAGATCCAGAGGTTG 8299
Qy 2041 CCAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100
Db 8300 CCAAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8359
Qy 2101 CGCTTCGAATTAAGGCGCGGAGCAGGAGTGTGCGGCTGCTGCTGCTGAGCTTGAATTC 2160
Db 8360 CGCTTCGAATTAAGGCGCGGAGCAGGAGTGTGCGGCTGCTGCTGCTGAGCTTGAATTC 8419
Qy 2161 TCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
Db 8420 TCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8479
Qy 2221 TCGAGCTGCGGCTGCTGCGGCGATGAGGCTTGAAGAGGATGATGAGGATGAGGATG 2280
Db 8480 TCGAGCTGCGGCTGCTGCGGCGATGAGGCTTGAAGAGGATGATGAGGATGAGGATG 8539

PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 4765
LENGTH: 4563
TYPE: DNA
ORGANISM: *Pseudomonas aeruginosa*
US-09-252-991A-4765

Query Match 12.3%; Score 520.6; DB 4; Length 4563;
Best Local Similarity 52.0%; Pred. No. 3,5e-105;
Matches 1511; Conservative 0; Mismatches 1329; Indels 63; Gaps 13;

1 ATGAGCATCAATCACTTGAACAGAGCTCGAGCAACAGGATATCAAGCTGGCGGCGAT 60
124 ATGAGCCTCGGGAATCTGTGGAACCTGCCAGCGCGCGCATGAACTCTGAGCGAG 183
61 GGGAGCGCTTCAGATACAGGCCCCCAAGAACGCTTGAACCGGAACCTGCTCTGA 120
184 GGGGCGCGCTGCTATGCGCGCCCCCGAGGCGCGCTCGACGCGCGCTCGCGAGCG- 242
121 ATCTCCGAGCAAAAGACGATCTTGAGATGCTCCGTCAGAGACTCCCGCAGAAATCC 180
243 --CTGCGGCGCGAGCGCGAGGCGCTGCTGGAACACCTGGAAGCGCGCGCTGCGCGC 300
181 ATCTGCGCGCGCGCGAGCGCGAGCGCTCGCTTCTCTCAAGACATCCAGAAATCC 240
301 GCGGAACCGGACTGCG 354
241 TACTGCTGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 297
355 TACGCTGCTGCG 414
298 GAATGCACTGAGCG 354
415 GAGTACGACTGCG 474
355 GTGCGCGCGCGCGAGCATGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
475 GTGCG 534
415 CCTAAGTGCAGCG 471
535 CCGGAGTGCAGCG 594
472 CCGGAGTGCAGCG 531
595 TTCCAGGCG 654
532 CGCGCTGCG 591
655 CAGTGGCG 711
592 CTGAGTATGATCTCATTAACGTTGACCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 651
712 TGCTGCGTGTGATTTACCTGTGCGACTGACGCGCGCGCGCGCGCGCGCGCGCGCG 771
652 CTGAGCTTCTAAGAAAGATTCGAGAGCTTCTCCCTGCTGAGCTCTCTGACCGCGAT 711
772 CG 831
712 TATGACTGCG 771
832 TATGTCG 891
772 TGAAGCG 831
892 TGGCTGCG 951
832 TCTACCTGGAAGAAATCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 891
952 GACAGCG 1008

892 GATCAGTAAAGCT 951
1009 CAGGCGCGCTCG 1068
952 GCATTTTCCGAGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1011
1069 GCTTCG 1128
1012 CTCTTCAACCG 1071
1129 GTACTCAACCG 1188
1072 ATGCTCTCTGATGATGACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1131
1189 CTGAGCTGCTGAGGATGAGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1248
1132 ATTCAAGCG 1191
1249 ATCGCGAGCAGATGTTGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1308
1192 CGAGAGCG 1251
1309 CGCGAATGCG 1365
1252 AGCGCGCTTAAACGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1311
1366 AGTGGCATGCG 1425
1312 TACACGAGCG 1371
1426 TACATGATCAGCG 1485
1372 GACCTGCTGCG 1431
1486 GACCTGAGATGCG 1545
1432 ATGCTGAGAGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1491
1546 ATGTTTGAAGCATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1605
1492 GTGCGCTTGTGCTTCCG 1551
1606 GACGCGATGCG 1647
1552 GCGCTGCGAGCG 1611
1648 TTGCGCGTATGCG 1707
1612 ATGAGCTGCG 1671
1708 GAGCGCGAGGATGCG 1767
1672 TCGCGCGAGCTTGGCG 1731
1768 GCCACCG 1827
1732 GTGCTGAGGAGAAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1791
1828 GTGATGCTGCG 1887
1792 GCGGCTGAGTGCAGATGCG 1851
1888 GCGGCTGATGTCG 1947
1852 CATGCTGAGTAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1911
1948 AGCGCGCGA-----GT 1996
1912 CCGGAGATCCAGCG 1971
1997 TCGGCTGCG 2052
1972 CCGATGATGCCATTCAGACACTTTCGATCTGCGGTATGTATCTACCTCGGAGATCC 2031

Db	2053	CCGCCCGGGGGAGGTGGCCGCGAGCACTTCGCTACAGTACTACCTCCGGCTCC	2112
Oy	2032	ACAGGGTTGCCAAGGGGGTGAATGATCATGGGGTCCGTCAACCATCTGGAC	2091
Db	2113	ACCGGCACGCAAAAGGGCGTGAATGCTCCAGCCATGGGGGTGAGCAACACCTCTCGAC	2172
Oy	2092	ATCAACGAGCGCTTCGAAATAGGGGCCGAGACAGGGGTCTGGGGCTCTCGCTGAGC	2151
Db	2173	ATCAACACCGCTTAGGGCGTGAAGCCACACACCGCTCTCGGACTTCGCGAGCTGAGC	2232
Oy	2152	TTGCATCTCTCGCTCTATGATGTGTTCGGATCTCTGGGCGGGCGGATCGATGTGGT	2211
Db	2233	TTGCAGCTCTCGATCTACGACTTCTTGGCGCCACCGCGGGGGGGCCAGGTGTCTC	2292
Oy	2212	CCGACCGGCTCCAACTGCGGCAATCCGGCGCATTTGGGCAGATTGATCGAACGAGACG	2271
Db	2293	CCGAGCCCGGGGGCGGAGGATTCATATGCACTGGGGGGAACGTGTGAACGCGACGCGC	2352
Oy	2272	GTGACGGTGTGAATCTCGGTCCGGCGCTGATGCGGATGCTCTGTCGACATTTTGAAGGT	2331
Db	2353	ATCACCCCTGTGMAACTCGGTCCGGCCCAAGCCAGATGCTCATCATTACTTGAGAGC	2412
Oy	2332	CGCCCCGATTCGCTCGCTAGGTCTCTGCGGCTTTGCTGCTGAGCGGCGACTGATCCG	2391
Db	2413	GAGCCGCAACGCTACCTCCGGGACCGCGCTGCGGTCTGTGGTCCGAGTATTCGG	2472
Oy	2392	GTGGGCTTCGCTGGGAGACTCCAGGCCATCAAGCCCGCGCTGTGGGATCAAGCTGGGC	2451
Db	2473	GTCAGCTTCGCGCAACCGATGGTGTGGGGCTGGGCGGACAGCGGCTGTAGCTTGGGC	2532
Oy	2452	GGGGCCACCGAAGCGTCGATCTGGCTCAATCGGGTACCCGTCGAGAAAGTCGACTATCG	2511
Db	2533	GGCGGCACCGAGGGGGCGATCTGGTCATTCAGAGAGCCGATCCCGCCGACGACCGAG	2592
Oy	2512	TGGCGAGCATTCCTTACGAGCCGTCCGTCGCAACGACGTTCCATGCTCTGATGAG	2571
Db	2593	CTGGGCACATCTCTTATAGCCGTGCCCTCGCGCGGCGAGACCTGGAAAGTCTTGATGCC	2652
Oy	2572	GCGCTCGAACCGCGCGCGGTCTGGGTTCCGGGGCAACTCTATTTGGCGGGGTGGGCTG	2631
Db	2653	CGCGGGCGGCGCTGCCCGCGCGGTGCGCGGCGAGATTCATATCGGGGGGTGGGCTTG	2712
Oy	2632	GCACTGGGCTACTGGCGGATGAAGAGAAAGCGCGCAAGACTTCTCTGTGCAACCCGAG	2691
Db	2713	GCGCTCGGCTAGCGCGGCGATCCGAGGCAACCGCGGAACGTTCTGTCCGACCCCGA -	2771
Oy	2692	ACCGGGAGCGCTCTTACAAGACCGGCGATCTGGGCGCGCTCTCTGCCGATGGAACATC	2751
Db	2772	--TGGCCCTCTCTGTATCGACCGGCAACCTCGGCGCTTCTGGCGGAGGAGATC	2839
Oy	2752	GAGTTCAATGGGCGGTGAGAGCAACCAATATCAAGCTTGGCGGATACCGGTTGAGCTCGGG	2811
Db	2830	GAGTTCTCTCGGCGCGGAGAGACGACGAGTGAAGATTGGCGGCGACCGCATGAACTGGCC	2889
Oy	2812	GAAATCGAGAAACGCTCAAGTTCGATTCGAACGTACGCGACGCGGTGATTTGTCCCGTC	2871
Db	2890	GAGCTGAGCGCGCGCTGTGCGCTCATCTCGAGGTCAATCTGGCGGACCATGTTGTCTC	2949
Oy	2872	GGGAACGACCGCGGCAACAAAGCT 2894	
Db	2950	GCGGAGCCACGAGCGGAGCTT 2972	

RESULT 10
 US-09-252-991A-4675/c
 : Sequence 4675, Application US/09252991A
 : Patent No. 6551785
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Mare J. Rubenfield et al.
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 : TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: 107196.136

```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 3142
; SEQ ID NO 4675
; LENGTH: 2865
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-4675

```

Query March	11.6%	Score 491	DB 4	Length 2865
Best Local Similarity	54.8%	Pred. Nov. 9.9e-99		
Matches 1121	Conservative 0	Mismatches 885	Indels 39	Gaps 6
QY	850	CGCTTCCGCGCACACGAGCAATGACTCCGTCGACTCTCGGGTCGATTTGAAACGGCGT	909	
DB	2811	CGTTTCCGCGCACTTCCACGCGCGCTCGACGAGGCCCTCTGCGAGGGCTTCGCGCGCGC	2752	
QY	910	GTCCGGGAGCGCGGGCTGACCCCGACGGGGCTACTCTGCGTGAATTTTCGAGGTATC	969	
DB	2751	GCCGCGCAACCGGCTCGAGGCGTCCGCGGCTGCGCTTGGCGGCTTCGCGCAACATC	2652	
QY	970	GGGCGCTGAGACGGAGACCCCGGTTTACGCTCAACAAAGCTCTTCAACCGGCTCCC	1029	
DB	2691	GTGCGCTGAGACCGAGCACCGCGCTTCTGTCTCACTGACCGTACTCAACCGCGCGCG	2632	
QY	1030	GTCCATCCGCGCGGTGAACGATATACCGGGGACTTCACTGTCGATGTCCTCTGGAATC	1089	
DB	2631	CTGATCTCCGACGCTCGCGCGAGTCCGTGCTGTACTTCAACCGCGCTCACTGCTGGACGT	2572	
QY	1090	GACACCACTCGCGCAAGAGCTTCCGAAACGGCGCTAAAGGTAATTAAGAGACGCTGG	1149	
DB	2571	GACACCGCGCACCGCGGACAGTTTCTGTGAGCTGTCCGACCGATCGCGCGGACAGATTTTC	2512	
QY	1150	GAACCGATGATCACTGCGACGTAAAGCGGTATCGAGGTCAAGCGAGCGCCCGCGGCTC	1209	
DB	2511	GACACCTCGACCAACCGACCTTCAGCGCGGCTCACTGCTGCGCGAACTGGCGCG---C	2455	
QY	1210	CTGGGGAATCCAGAGGGGCAATTGTCCCGGTGTCTCAAGAGCGGCTTAAACACAA	1269	
DB	2454	CGGCGCGGTCCGCGCGCGCACTTATGCTCGGTGTCTTACCACTGAGCATTCGACGCTG	2395	
QY	1270	GTCGTGGTGTCACTCGTTGCGAGGCTCGGAATCTCGGTGTACACAGCAGCAGACT	1329	
DB	2394	CAGCGCGCTCTCGCGATGAGGAGGCGCGCGCGCCGACACGCTACATGATCAGCAAC	2335	
QY	1330	CCTCAAGCTGCTGTGATCATCACTCTTAAGAGACGATGGGACCTCTGCTCTCGGTG	1389	
DB	2334	CCGACAGGCTGTGGCTGAGACTGCGAGATCAACGACGATTCCGCGCGCTGAGGATGGCTGG	2275	
QY	1390	GACATCGTCAGCGAGAGTTCCTCCGCGCGACCTTCTGAGCAGCATCTCGAAGCTACGTC	1449	
DB	2274	GACGTACGCTCTCGGTTGTTCTCCGAGGAGCGCAGCGGAAAGCATGTTTCAAGATTGCTC	2215	
QY	1450	GTTTTTCTCCGCGCGCTCATCTAGAGAAACATAGGGGTAAAGGTGCGCTGTTCGCTTCG	1509	
DB	2214	GGGCTGTCTCCGCGCGCTCGAGCGCAAGCCGCGCGCTGAGCCAGCGGCTGATCCAGGAA	2155	
QY	1510	CCTGCGCCAGTAAAGCGCGGCGAGCGCAAAACGCAACCAACGCGCTGTGAGCGAGCAT	1569	
DB	2154	CCGCTCAAGGCGCGCGCGCGCGCTT-----GCCCGGTAGTGGCCCG 2113		
QY	1570	ACGCTGACGCGCTGTTCGCGCGCGGAGTGCAGCTGCCCATGCAAGCTGCGCGTGTG	1629	
DB	2112	AGCATTCGCGCGGTTTCGCGAGCTGTGCTCTGACCCCGACCGCACGCGCATCCAC	2053	
QY	1630	TCCGCGCGCAAGAGCTCACTAGAGAGTTTGGCGCGTTGCGGCGCACTTGGCGCG	1689	
DB	2052	GATGCGCGCGGACATGACTTACCGCCAGGTGCGCCAGACGCGCACCGCTCTGCGCGC	1993	

1690 CGGCTGCGGAGACGAGGCGGACGCGGAAACACATTGTCGCGGTGATGAGAAAGGC 1749
1992 GTCCGGAAGCGGACGCGGCGCGGCGCGGTGCGGTGATGTCGCGGAAAGGC 1933
1750 TGGAGCAGGTTGTGCGGTTTCCGCGGTGTCGAGTCAGGCGCGCTTACGTGCGGATC 1809
1932 GCCCGCAATTGTTGCGGATGATCGGCACTCCGAGGCGCGCGCTTATGTCGCGGTG 1873
1810 GATGCGGACCTTACGCGGAGGATTCACATCTACCTGATCATGTGAGGTAAAGTC 1869
1872 GACATCCGCGACCTTCCGTCGCGCGCGGAGGATCTCCGACGCGCGA-----A 1822
1870 GTGCTGACGACGACATGCTGATGAGCAAACTGTATGAGCGCGCGGATCCAGCGGCTG 1929
1821 GTGGTCGCGCTGTGCTGCTGGAAGCGA--TGTCCGGAAGTGTGCGGTGCGCGTGTG 1764
1930 CTCGAGACGAGGCGCGGCGGTGGAAGCGGACGCGGACCTTCGATGATGCGCATTCAG 1989
1763 CCATGACCGGCTG---GCCGCCGACAGCGCGCTGCGCGCAACCGCGCGCGAGGTG 1708
1990 AACCTTCGGAATCTCGGATGATCTTACATCACTCGGGAATCCAGAGGTTGCCAAGGG 2049
1707 GCGCGGACGACCTTCCCTGATGATCTACATCTCGGCTCCAGCGGACGCGCAAGGCG 1648
2050 GTGATGATGATCATCGGCGGTGCGGTCAACCATCTGACATCAAGCGGCTTCGA 2109
1647 GTGATGATGATGATGATGCGGCGGTGAGCAACGCTGCTGACATCAACAGGCTTACG 1588
2110 ATAGGCGCGGAGACAGGCGGTGTCGCGCTCTCTGCTGAGTTGATCTCTGCTGAT 2169
1587 GTGAGCGCAACGACCGGCTCTCGGCTCGCGGAGCTGAGCTGACCTCTCGCTTAC 1528
2170 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2229
1537 GACTTCTTCCGCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1468
2230 CGCGATCGGCGGATGAGGCGGAGATGATGATGATGATGATGATGATGATGATGATGAT 2289
1467 AGCGATCATGCACTGCGGCGGAGCTGCTGAGAACGCGCATCACTGTGGAACCTG 1408
2290 GTGCGGCGGCTGATGCGGATGCTGCTGAGCACTTTGAGGCGTCCCGGATTCGCTGCT 2349
1407 GTGCGGCGGCGGAGCGGATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1348
2350 AGGTCTGCGGCTTTCGCTGCTGAGCGGCGGAGCTGAGTCCGCGGCGGCTGCTGCGGAG 2409
1347 CGGGAACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1288
2410 CTCCAGGCGGATGAGCGGCGGCTGCTGCTGATGATGATGATGATGATGATGATGATGAT 2469
1287 TGTGCGGCGGCGGCTGCGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1228
2470 ATCTGCTCATGCGGATCCCGGTGAGAACTGCACTATGCTGCGGCGGAGCATTCCTTAC 2529
1227 ATCTGCTGATGAGCACTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1168
2530 GCGCGTCCGCTGCGGAGCACTGATGCTGCTGATGATGATGATGATGATGATGATGATGAT 2589
1167 GCGCGTCCGCTGCGGAGCGGAGCGGAGCTGAGATCTGAGATGCGGCGGCGGCGGCTGCG 1108
2590 GTCTGAGTTCCGCGGCACTTACATGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 2649
1107 CCGGCGGCTGCGGAGATCATTCGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1048
2650 GATGAGAGAGAGCGGCGGAGCTTCTCTGTCACCCCGAGACCGGAGCGGCTTAC 2709
1047 GATTCGAGCGGACCGCGGAGCTTCTCTGTCACCCCGA---TGGCGCTGCTGCTGAT 991
2710 AAGAGCGGAGATGCTGCGGCGGCTGATGCGGCGGAGTGAATCATGAGATCATGAGGCGG 2769
990 CGCAGCGGAGCTTGGCGGCTGATCTGCGGCGGAGATGAGATCTTCTGCGGCGGAG 931
2770 GACAACTCAAGCTTTCGCGGATACGCGCTTGAAGCTGCGGAGATGAGAGAAAGCTC 2829

930 GACGACGAGTGAAGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 871
2830 AAGTCGATCCGAGACGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2889
870 TGCCTCATCCGAGGTCATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 811
2890 AAGCT 2894
810 AGCCT 806
RESULT 11
US-09-252-991A-4815
Sequence 4815, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 4815
LENGTH: 2922
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4815
Query Match 10.7%; Score 455; DB 4; Length 2922;
Best Local Similarity 50.1%; Pred. No. 7.9e-91;
Matches 1352; Conservative 0; Mismatches 1295; Indels 50; Gaps 7;
156 CGGTGAGGAGTCCCGGAGATTCATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 215
264 CCGGCTGCGGCGGCGGAGCGGTCGCGGCTCCAGCGGCGGAGATCGATACGCGCT 323
216 TCTCTGACAGACATTCAGAGATCTTACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCT 272
324 CGAGCTGCTTCCGTCGAGCGGCTTACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 383
273 CCGGAGGAGATTCAGCGCTTATGCGGATGATGATGATGATGATGATGATGATGATGATGAT 332
384 CAAGTCAAGTCCATGCTTCTGGAATTCGCGGCGGCGGAGTTCGACCGGAGCGGCT 443
333 GAGCGCGGCTTTCGGAAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 392
444 GCGCGGCGGCGGAGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 503
393 CGACATGATGAGGATTCAGAGCTTAAAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 452
504 CGGTGCGGAGATCTTCCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 563
453 CGGCTGAGCGGAGCAACGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 512
564 CACTTGAAGGTGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 623
513 CCGGATCTATGACACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 572
624 TGAATGCTGCGGAGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 683
573 GCGGCAACCGGCTGCGGCTCAGTATGATCTCATTAAGCTTGAAGCTTGAAGCTTGAAGCTT 632
684 CGGCGAGGATGCTTCTGCTGATGATCTGACCTTCTGCGGCGGCGGCGGCGGCGGCGGCGG 743
633 CATCATCTTCAAGGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 692
744 CCGCTGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 803

QY 693 GAGCTCTGTACCGGATTATGTAATCTCCGCTGAGATCTCCGAGAAATCTGAGCGCA 752
DB 804 GCGCGTCACTTTCGCGACTA---CTTGCGCAACCTGTGCGGCAACGCGCGAGCGCC 860
QY 753 TCAGCATGATGATTACTTGAAGGCGGCAATCGCGAGCTCCCACTCCGCGCAAGCT 812
DB 861 GCGCGGCGCGCGGACTTACTGTGCTGGAAGGCTTGGCGCGCTTGGCGAGCGCGCGCTT 920
QY 813 TCCGATGAAGCGCGATCACTTACCTGAAGAGATCGCTTCCGCGACACGAGCAATG 872
DB 921 GCGGTGGCTGGCGCGCGAGAAAGATCCGCGACCGCGCACCGCGCGCTGGCAATTCA 980
QY 873 GCTGCGGTGGAATCTCTGGGGTGTGATTTGAAGGCGGCTGTGGGAGAGCGGGCTGACCC 932
DB 981 GCTTTCGCGCGGAGAGCGCGGCGCTGAGGCTTTCGCGCGACATGAGCGTGAACCTT 1040
QY 933 GACGGGCTGATCTCTGGTGCATTTTCCGAGGTGATCGGGCGCTGAGCGCGAGCCCG 992
DB 1041 GTTCAGCGTGTGTGCGTGGCGCTTTCGCGCTGTGTCTGGCGCTGAGCGAGAAAGCGCGA 1100
QY 993 GTTTACGCTCAACATTAACCTCTTCAACCGGCTCCCGTCCATCCGCGGTGAACGATAT 1052
DB 1101 ATTCTCTCAACGTCGCTGTTGTCATGCGGATGCGAGACCGCGATCGGCGAGGT 1160
QY 1053 CACCGGGGACTTACGTCGATGTCTCTCTGAGCATGACACCACTCGCGACAAAGCTT 1112
DB 1161 GATCGCGCACTTCAACCACTTGTGTGCTGAGTCCGAGTGCAGAGCGGGGTGTCTT 1220
QY 1113 CGAAGCGCGCTAAGCGTATTTCAAGAGAGCTGTGGGAAGCGATGATCACTGCGAGT 1172
DB 1221 CCGCGAGGGGTGAAGACTTCCAGCGCACTTCCAGAGCATGACACGCGCGAT 1280
QY 1173 AAGCGTATCGAGTTCAGCGAGAGCGCGCGCGGTCTGGAGATCAACGAGCGCAT 1232
DB 1281 CCGCGCGCTGAGGTCTCCGCGAGCGCGCGCG---CAGGGCGACACGAGCTCGAG 1335
QY 1233 GTTCCCGGTGTGTCTACAGAGCGCGCTTAAACGAGCATGTGTGTGATCACTCTGTGA 1292
DB 1336 GCGGTGTGTGTCTACAGCAACTGTGGCGAGAGGCTTGTCTCCGCGCGCTTCCGCA 1395
QY 1293 GAGGCTCGGAACCTCCGCTGTACACAGAGCGAGACTTCTCAGCTCTGTGATCATCA 1352
DB 1396 CGCTTTCGCGGATCTCCAGCACTGTCTGTGAGACCGCGAGGTCTGGCTGACCAACA 1455
QY 1353 GCTCTTACAGACGATGAGGACCTGTCTCTCGGTGAGACATGTGACGAGGTGTCC 1412
DB 1456 GCTCTACGCGGTGGGCGAGCGGTATCTGTGTGCGCTGGGACAGGCTGTGCGCTGTCC 1515
QY 1413 GCCCGACTTTCGAGAGCATGTCTGAGAGCTATGCTTCTCCGCGCGCTCACTGA 1472
DB 1516 CGAAGGCTGTGCGGAAACATGTTTCAAGCTTACGTGGGCTGTCTCAAGGTCTTGTGA 1575
QY 1473 GGAACATGAGGAGTGAACAGGTGCGCTGTGTCTTCCGCTGCGCAGCTAGAAAGCGGGC 1532
DB 1576 CAGGCTGTGGAGCAACCGCGCATGTGCGTGTGCGCTGGCGGACAGCGCGCGGGC 1635
QY 1533 GAGCGAAACGAGCAACGCGCTGTGAGCGAGCATAGCTGTGACGCGCTGTTCGCGC 1592
DB 1636 CTTGCTCAACGAGCAACCGCGCATGCGCACGCGCGCACCTGTCATCGGCACTTCTTCT 1695
QY 1593 GCGGGTGTGAGAGCTCCATGAGAGCTGCGCGGTGTGTGCGCGCGGAGAGCGTCACTGA 1652
DB 1696 TCGCGCGCGCGAGCGCGCGAGTCCGACGCGGTGTCTTATCGGACCAACGTGTCAACCG 1755
QY 1653 CGAAGAGCTTTCGCGCTTTCGCGGAGCTTTCGCGCGCGGCTGCGGAGAGCGGGCAGG 1712
DB 1756 CCGGCACTGTGCGGAGGTGTGCTGTGCGATCGCGCGCGCTGCGGAGAGCGGGGTGG 1815
QY 1713 CCGGAACATTTGTGTGCGGTGTGTATGAGAAAGCTGTGAGAGAGGTGTGCGGTCT 1772
DB 1816 CCTGTGCGAGCGGTGTGAGGTGTGAGCTGTGCGCGCGGACCGGACAGGTGTGCGGTAT 1875

QY 1773 CCGGTGTGTGAGTACGCGCGCGCTTACGTGCGATGATGCGGACCTAACCGGCGAGCG 1832
DB 1876 CCGGTGTGTGCGGAGCGCGCTGTGATGTGCGGTGACATGACAGCGCGCGCACG 1935
QY 1833 TATCCACTACCTCTCGATCAATGTGTGAAGTCTGTGTGACGACGCAATGCTTGA 1892
DB 1936 GCGGCGCTGATTCGAAGAGCGCGCGGGGTATGTCTGTGATCAACGAGAGAGCATTC 1995
QY 1893 TGGCAACTGTATGTGCGCGCGGAGATTCAGCGGTGTCTGTGAGCGAGCGCGGTGA 1952
DB 1996 GCGGCTGTGTGCGCGCGCTGTGATGTGACGCGCTGTGTGCGCGCGCGCGGTGGC 2055
QY 1953 AGGCGAGCGGACCAAGCTTCGATGATGCCATTCAGACACTTGTGATCTGCGTATGT 2012
DB 2056 CCGGTGTGCGGTGCGCGCGCGAGCGA-----GTGCTTATGT 2091
QY 2013 CATTTACACTTCGAGATCAGAGGTTTCCAGAGGGGTGATGATCATTCGGGCTGC 2072
DB 2092 GATTTACACTTCGAGCTCACCGGGGTGTCAGAGGCTGAGGTCAAGCACGCGCGCGC 2151
QY 2073 GGTCAACCACTCTGACATCAACGAGCGCTTGAATAGGCGCGGAGACAGGTGTCT 2132
DB 2152 GATCAATACATTCAGCGCGCTGTCTGACCTGTGCGGTGAACGATCGATGCTTCT 2211
QY 2133 GCGGCTCTCTGAGCTTGTGATCTGTGCTATGATGTGTGAGATCTGCGGCGC 2192
DB 2212 GCGGCTCTCTGAGCTTGTGATCTGTGCTGTGCTTGTGACCTGTGTGCGGCGCTGCGC 2271
QY 2193 GCGGCTACATTCGTGTGTGCGGACGCTGTCAACGCTGTGCGATTCGCGGCAATGGGCA 2252
DB 2272 CCGGTGCAAGCTGTGTGTGCGCGCGCGGAGCAAGCGCGCGCATGTGCTGTGCGGCGA 2331
QY 2253 GTTATTCAGAGAGAGAGGTGACGCTGTGGAATCTGTGTGCGCGCGCTGATGCGATCT 2312
DB 2332 GGTATTCAGAGCGGCTGTGTGTGAGCTGTGTGAACCTGTGTGCGCGCGCTTGTGAGATGCG 2391
QY 2313 GGTGAGCATTTTGAAGGTGCGCGCGATTCGCTGTGATGTCTGCGGCTTGTGCTGT 2372
DB 2332 CTTGAGCTTGTGCGGAGCGAGCGAGCGA-----CATGTGAGCTGTGCGGCGGTGTCT 2445
QY 2373 GAGCGGCACTGTGATTCGCGGTGTGCGCTGTGCGAGCTGTGAGGCAATCAGGCGC--G 2429
DB 2446 GTTCGCGCACTGTGTGTGCGCGCGCTGTGAGCTGTGCGCGCGCTGTGCGCGCAAG 2505
QY 2430 CGTGTGATGATCACTGTGTGTGCGCGCGCAACGAGAGCTGTGATCTGTGTGATGCTTCC 2489
DB 2506 CTGCGCGCTGTGATGTGTGTGTGTGCGCTTACGAAACGAGCAATGTGTGAACCTGTGAG 2565
QY 2490 CGTGAAGAACGTGACCTATCTGTGTGTGAGCATTCCTTACGAGCGCTGTGCGCAACA 2549
DB 2566 CTTGATGATGCGCGCGCGCGCGCTGTGATTCCTTACGCGCGCGCATTTCCCGGACA 2625
QY 2550 GAGCTTCACTGTGTGATGAGGCGCTGAACCGCGCGCGGTGTGTGCGGCGCAACT 2609
DB 2626 GGTCTACGCGGTGTGTGACACCAAGCGCGCGGACGCTGTGCGGAGCTGTGTGCGGAGCT 2685
QY 2610 CTACATTTGCGGCGGTGTGCGCGCGCTGTGCGGCGGATGAGAGAGAGAGCGCGCA 2669
DB 2686 GTGATGTGCGCGCGCGCGCGCGCGGTGTGTGCAACGATTCGCAACTGAGCGCGC 2745
QY 2670 GAGCTTCTGTGTGACCGCGAGACGCGGAGCGCTTACAGAGCGCGGATCTTGGGCG 2729
DB 2746 GCGCTTGTGT-----CAGAGAGCGCGCGCGGTGTGTGTGCGAGCTGTGCGGCG 2799
QY 2730 CTACCTGCGCGATGAAACATGAGTTTATGTGTGTGAGAGCAACCAATCAAGCTTTCG 2789
DB 2800 CTACGTGCGCGAGCGGATCCCTGAAATCTGTGTGTGTGCGAGCAGAGGTGAAGTGTG 2859
QY 2790 CGGATACGCGCTGTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2846
DB 2860 CCGCGAGCGCATGT 2916

RESULT 12
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 8.5%; Score 358.2; DB 3; Length 4411529;
Best Local Similarity 48.4%; Pred. No. 7.2e-69;
Matches 1252; Conservative 0; Mismatches 1283; Indels 54; Gaps 7;

```
QY 286 CACGCTATCCGAAATGAGCTGTACGATCTCGACGTCCGAGCGCTGACGCCGCTTT 345
DB 2675453 CACCTCTACGTGAGTTCATGTGTCGCGGATCCGAGACAGCTCCGCGCGGCC 2675404

QY 346 CGAAAGTCTGTCGCGCGGACGACATGCTTCGCGGCCACACGCTCCGACATGATGAG 405
DB 2675403 ACCCGGTGGGCGCTCGGAGACCCGATGCTCGGGGTGAGTTCTTCCGACGCCAG 2675344

QY 406 GTGATCGAGCTTAAAGTGCACGCCGACATCGAGATCATGATCTCGCGGGCTGCACCG 465
DB 2675343 CGCATTCGCGCGCGCGGAGTCTGCGACCTTCCATCATGATGTGCGACATCGCTAC 2675284

QY 466 AGCAACGCGGAAAGGAGGCTGCTGTCGTCGAGATGCGATGCCACCGCATATATAC 525
DB 2675283 GTGGCCCGGATGTGTCGATCAGCGGTTGCGGGGATTCGCGACCCAAATGCAACG 2675224

QY 526 ACCGAGCCCTCCGCTCTATCAGCTGCTGCGCTTCGCTGGAAGCGGCAACCGCT 585
DB 2675223 CAGCTCGAGCGGTGCGGTATTCGACTTGTGATGATTAACCGGGGAGCGCACCCGC 2675164

QY 586 CTGCTGCTCAGTATGATCTCATTTAACTTGAAGCTTGAAGCGCTGTCCATCTTTCAAG 645
DB 2675163 CTACATGTGCACTGCAATGAGGCGCGCTGACGCGATGAGCTACCGCATTTGCTGCGC 2675104

QY 646 GACTGCGTCAAGCTTCTAAGAAATATCCGAGACTCTCTCCCTGCTCTGAGCTCTCGTAC 705
DB 2675103 GACTGCGCGGCTCTTATGACGCGCTGTAACCGCGGACATGCGGCTTACACTTACCGGAA 2675044

QY 706 CGCGATTATGTAATCTGCGCGTGAAGTCTGCAAGAAAGTCTGAGCGCATTAAGATGATG 765
DB 2675043 TACCGGAGGCTATGAGGCGGAGAGACGCTGCCCAACCGCTTCGCGACCGCAACCG 2674984

QY 766 GATTACTGGAAGCGGCGCATTCGCGAGCTCCCACTCCGCGGACGCTTTCGATGAAGCC 825
DB 2674983 GACTGTGCGGCGCAACGCAATCCGCGAGCTGCGGATTCGCGCGCTGCTACCGGCGC 2674924

QY 826 GATCATCTACCCCTGAAGAGATTCGCGTTCCGCGACACGAGAACATGCTGCGCTGAGAC 885
DB 2674923 GCGCGGAAACGAGACCGGCGCGGACGACACCGCGCC--TGGCACTGTGCTAGACCTCGAG 2674867

QY 886 TCTCTGGGTGATTAAGCGCGGTCTCGGGAGCGCGGCTGATACCCGACGCGGCTCATC 945
DB 2674866 ACCCGGACGCGGTCTTGGCCCGGACCGGGCGCGCGGATACCCCGCGCATGACGCTG 2674807

QY 946 CTGGCTGATTTTCCGAGGTGATCGGGCGCTGGAAGCGGAGCCCGGCTTACGCTCAAC 1005
DB 2673744
```

```
DB 2674806 GCCCGGCTTTGCCAAGTGTGCGCGGTGTCAGGCTGTGCGGTTCTCTGTGAC 2674747
QY 1006 ATTAAGCTTTCAACCGGCTCCCGTCCATCCGCGGTAAGATATACCGGAGACTTC 1065
DB 2674746 CTGCGGTGTTTCACTGCGGACGAGCCCTGATCATCAAGCTGCTGTGTGATCTTC 2674687
QY 1066 AGCTGATGATCTCTCTGAGATGACACCACTGCGGACAAAGAGTTTGAACGCGCT 1125
DB 2674686 ACTCTCGCTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2674627
QY 1126 AAGCTATTCAGAGACGCTGTGGAGAGATGATGATGATGATGATGATGATGATGATGAT 1185
DB 2674626 CAGGCGGACGAGAAAGCCCTCGAGCGCCGAGCGCCACAGTATACCCGCGCTGTCT 2674567
QY 1186 GTCCAGGAGAGGCGCGCGGCTCTGAGGATATCCAAAGGCGCATTTGCTCCCGTGTG 1245
DB 2674566 GTGCTGCTGATCTTACGCGCCGACCGTGGACCCAGGTGCTGACCGGTGATTTAC 2674507
QY 1246 CTACAGAGCGGCTTAAACAGCAAGTCTGTGTGTACCTGCTGACAGGCTCGGAACT 1305
DB 2674506 AGCGCGTGGGCTCTCGGAGCTTCTTCTGCGCGGACGTCACCGAGCA--ATTGCGACA 2674450
QY 1306 CCGGTATCACGACGACGACATCTCTAGCTGCTGCTGATCATCTAGCTCTTACAGAC 1365
DB 2674449 CCGGATGATCATTTTCCGAGGCGCCGAGGTGCTGCTGACGCGCAGTCAACGAGTTT 2674390
QY 1366 GATGGGACCTGCTCTGCGGAGGACATGCTGACGAGAGTGTTCGCGCGGACCTTCTG 1425
DB 2674389 GACGCGGTGTGCTGTGATATGAGATGTCCGAGGAGGCTTTTGACACCGGCTCATC 2674330
QY 1426 GACGATGCTTGAAGCTTACGTCGTTTCTCGGCGGCT-----CACTGAGAACCA 1479
DB 2674329 GACGCTATGTTTACCCACAGAGTGAAGATTTGCTCGGTTGCGCGGAGGAGACGCGC 2674270
QY 1480 TGGGTGAACAGGTGCGCTGTTCTGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1539
DB 2674269 TGGATGCGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2674210
QY 1540 AACGCGAACCAAGGCGCTGTGAGGAGATAGGCTGACGCGGCTGTTCGCGCGGCGGTC 1599
DB 2674209 AACGTCGACCGCGCGCGCCGACGAGCGGCTGACGAGGCTTTTTCGCGCAGGCC 2674150
QY 1600 GAGGAGCTGCCATGACGCTGCGGTGTGTGCGGCGGCAAGACGCTTACGATCAAGAG 1659
DB 2674149 CAACAGAGCGCGGAGCGCGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2674090
QY 1660 CTTCGCGCGCTTGGCGGCACTTGGCGCGCGCTGCGCGGACGAGGCGGACGCGCGGAC 1719
DB 2674089 CTGCGCGACAGGATCGCGGCTGCGCGCGCTGCTGCTGCGCGGCTTACGATCGCGC 2674030
QY 1720 ACATTGTGCGGCTGTGATGAGAAAGCTGTGGAAGAGCTGTGCGGCTTCTGCGGCTG 1779
DB 2674029 GACACGTCGCGGTGCTGAGTCCGAAAGCGGCGCAAGAGGCGGCTGTCTGCGGAT 2673970
QY 1780 CTGAGTACGCGCGGCTTACGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1839
DB 2673969 TTGGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2673910
QY 1840 TACTCTCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1899
DB 2673909 CGCATCTCGGAGACCGGCTTCTGATCACTTACCGCT-----GTTTTCGCGC 2673865
QY 1900 CTGTGATGCGCGCGGAGATCCACGCGCTGCTGAGCGAGGCGCGGCTGAGAGGCGAC 1959
DB 2673864 CCGCATGCGCAAGTGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2673805
QY 1960 GCGACACGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2019
DB 2673804 GCGCGGAGAAATTCGTCGCGGCGCTGAGGATCCACCGGCTGCTGCTGCTGCTGCTGCT 2673745
QY 2020 ACTTCGAGATCAACAGGCTTCCCAAGGAGGATGATGATGATGATGATGATGATGATGATGAT 2079
DB 2673744 ACTTCGAGCTTACCGGAGAACCCCAAGGATGATGATGATGATGATGATGATGATGATGAT 2673685
```


Db 2671446 CCAACAGACGCCGACGCGCGCGGTTCGCGACGTTCCGGCGACTTGAAGTACGCCA 2671387
QY GCTTTCGCGCGGTTTCGCGCGGCACTTGGCGCGCGGCTGCGCGAGCGAGGGGACGCGCCGA 1718
Db 2671386 ACTGCGGCAACGAGGATGCGCGGTGCGCGCGCGCTGCGTGTGCGGGGCTTACGAGTGG 2671327
QY 1719 CACATTGGTTCGCGGTGTGATGAGAAAAGCTGGAGAGGTTGTTCGCGGTTCGCGGT 1778
Db 2671326 CGACACCGTCCGCGGTGTGAGTCCGAAAAAGGCGGACAAAGTGGCGGCTGTCTGGGAT 2671267
QY 1779 GCTCGAGTCAGCGCGGCTTACGTGCGGATCGATCGATCCGACCTTACCGGCGAGCTATCCA 1838
Db 2671266 TTGGCGCGCGGCGGTTCATCTGCGGCTTACGCGGCTTACCGGCGGCGGCGG 2671207
QY 1839 CTACTCTCTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1898
Db 2671206 GCGCATCTTGGCGAGCGGTTGCGTCACTT-----AGCGCTCTTTTGGCG 2671162
QY 1899 ACTGTGATGCGCGCGCGGATTCAGCGGCTGCTGCTGAGCGAGCGCGGCTGAGAGCGA 1958
Db 2671161 CCGCGCATGCAAGTGGCGGATGCGGCTCCGACGCTGTTGCTGCTGAGCGTCTTGGCGG 2671102
QY 1959 CGGCGACCAAGCTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2018
Db 2671101 CGCGCGCGGAGATTGTCGCGCGGCTTACGCGATCCACCGCGCTGCGCTTATGTGTGT 2671042
QY 2019 CACCTCGGAGTTCACAGGATGCGGATGCGGAGGAGTGTATGATGATGATGATGATGATGAT 2078
Db 2671041 CACCTCGGAGTTCACAGGATGCGGAGGAGTGTATGATGATGATGATGATGATGATGAT 2670982
QY 2079 CACCATCTGATCATCAAGCGGCTTCAAAATAGGCGCGGAGACAGGATGCTGCGCT 2138
Db 2670981 CACCGTGAAGACTTTCATCTCGGCACTTCAAGTGAAGCGCGGAGATGCTGCGCTTGGCT 2670922
QY 2139 CTCCTCGTGAAGTTCATCTCGGCTTATGATGATGATGATGATGATGATGATGATGAT 2198
Db 2670921 GCGGACGCTGAGTGGCGCATGTGCTGCTGAGCACTTTCGCGCGCTGCTGCGGCG 2670862
QY 2199 TACGATCTGATGCGGAGCGGCTTCAAGCTGCGGATGCGGAGTATGATGATGATGAT 2258
Db 2670861 AGCGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2670802
QY 2259 CGAAGAGAGAGTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2318
Db 2670801 CGACACTTACGAGGTCACGCGGTTGATTTATGCTGCGGATGATGATGATGATGATGAT 2670742
QY 2319 GCATTTTAAAGGTGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2378
Db 2670741 AGTCGCGCGGAGCGCGCTG-----TCGTCGCTGCGAGCGGATGCTGCGGCG 2670694
QY 2379 CGACTGATGATGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2438
Db 2670693 CGACTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2670634
QY 2439 GATGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2498
Db 2670633 CGCGGAGTGTGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2670574
QY 2499 CGTGAC-----CTATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2549
Db 2670573 CGCGGCGGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2670514
QY 2550 GACGTTTCAAGTGTGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2609
Db 2670513 CGCTGCGGATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2670454
QY 2610 CTACATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2669
Db 2670453 GTGGGTGTGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2670394
QY 2670 GAGCTTCTCTGTCACCGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2729
Db 2670393 GCGCTTCTGTCAGTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2670340

QY 2730 CTACTGCGCGGATGAGAAACATGATGATGATGATGATGATGATGATGATGATGATGAT 2789
Db 2670339 CTACTGCGGAGACGCGGACCGCTGAGTGTGCTGCGCGGCTGCGGATGATGATGATGAT 2670280
QY 2790 CGGATCCGCGGTTTAACTCGGCGGAAATGAGAGAAAGCTCAAGTGCATCCGAACTGAC 2849
Db 2670279 CGGATCCGCGGTTTAACTCGGCGGAAATGAGAGAAAGCTGAGCGGCTGCGGCTGCA 2670220
QY 2850 CGAGCGGCTG 2859
Db 2670219 CGCGCGGCGG 2670210

RESULT 14
US-09-252-991A-4678
; Sequence 4678, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252, 991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4678
; LENGTH: 1842
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4678

Query Match 8.2%; Score 348.6; Db 4; Length 1842;
Best Local Similarity 51.7%; Pred. No. 1.6e-67;
Matches 946; Conservative 0; Mismatches 839; Indels 44; Gaps 5;

QY 1064 TCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1123
Db 1842 TCACACACCTGTTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1783
QY 1124 CTAGCGATTCAGAGAGAGTGTGAGAGAGATGATGATGATGATGATGATGATGATGATGAT 1183
Db 1782 TGAAGAGTTTCAAGCGCACTTCAAGAGATGATGATGATGATGATGATGATGATGATGAT 1723
QY 1184 AGTTCAG 1243
Db 1722 AGTTCAG 1668
QY 1244 TGTTCAG 1303
Db 1667 TCGCAG 1608
QY 1304 CTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1363
Db 1607 ATTCACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1548
QY 1364 AGATGAG 1423
Db 1547 TGGGCGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1488
QY 1424 TGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1483
Db 1487 CGGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1428
QY 1484 GTGAAG 1543
Db 1427 AGAGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1368
QY 1544 CGACCAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1603

Db	1367	GCACGCGGGCATTGGCCCAATGGGCGGCACCTTGCATTCGGACTTCTTCCCTTGGCGCGCGC	1368
OY	1604	AGCTGCGCATGCACTCGCCCGTGGTGTGCGCGCGCAACGCTCACTGAACAGACTTT	1603
Db	1307	AGGCGCCCGAATGCCGACCGCGCTGCTCTTATCCGCATCAACCTGTCCACCCGCGCGCAACTGG	1248
OY	1664	CGCGCGCTTCCGCGCGCACTTGGACGCGCGGCTGCGCGACAGAGGGGCAAGCCCGAAACAT	1723
Db	1247	CCGACGCTGCGCTGGGCATTCGCGCGGCGCGCTGCGCGAAGCCGGGTGGCCCTTGCGACAG	1188
OY	1724	TGCTGTCGCGTGGATGATGAGAAAGGCTGGAGACAGCTTGTGCGCGCTTCTCGCGTGTCTG	1783
Db	1187	CGGTTCGAGGTGACCTTCGCGCGCGGACCGCAGACAGGTGCGCGCGGTATTTCGCGCTGTGCG	1128
OY	1784	AGTACGAGGCGCGGCTTACGTCGCGCATTCGATTCGACCTTACCGCGGAGGGTATTCACCTTAC	1843
Db	1127	CCGACGCGCGCTGCTGCTACGCTGCGCGCTGCGACATGACACAGCCGCGCGACGCGCGCTGA	1068
OY	1844	TCTTCGATCATGTGTGAGGTAAAGCTCTGTCTGAACGACGCAATGCTGTGATGGCAAACTGT	1903
Db	1067	TGCAAGAGGCGCGCGGGGTATGCTTCGGGATACCGAGAGGACGATCCGACGCGCTTGC	1008
OY	1904	CATGCGCGCGCGGATTCAGCGCGCGCTGCTCTGTGACGAGGCGCGGCTCGAAGCGCACGGCG	1963
Db	1007	GCGCGCGCTGATGTTCAGCGCGCTGTGCTGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCG	954
OY	1964	ACCAAGCTCCGATGATGATGCGCATTCAGACACTTCGCGATTCGCGCTATCTCATCTTACCT	2023
Db	953	-----TGCGCGTGGGCGCGCGGCGGCGAGCGAGTGTCTATGTATCTTACCT	912
OY	2024	CGGATTCACAGGCTTCCCAAGGGGGTGAATGATGATCATCGGGGTGCGTCAACACCA	2083
Db	911	CGGGCTCCACCGGGGTGCGCAAGGGCGCTGAGGTACGACACGCGCGCGCATCAATACCA	852
OY	2084	TCTTGGACATCAACAGCGCGCTTCCGAATATAGGGCGCGGAGACAGGGGTGCTGGCGCTCTCT	2143
Db	851	TGCAAGCGCGCTGTGCACCTGCTGCGGGTGAACGATCGGATCGCTTGGCGCTTGGCGCTTCCG	792
OY	2144	CGTGAAGCTTCGATCTCTCGGTCTATATGATGTGTTCGGGATCCTTGGCGCGCGCGCGTACGA	2203
Db	791	CGCTTGAATTGATCTGTCTGTGCTTTCGACTGTGTCGGCGCGCTCGGCGCGCGTGCACCC	732
OY	2204	TCTGCGTCCCGGACCGCGTTCGAAGCTGCGCGCATTCGCGCGCATTTGGCGCAAGTTGATTCGAC	2263
Db	731	TGTCTCTCCCGCGCCAGGAACAGCGCGCGGATGCGCTGTGCGCGGAGGCTATTCACGC	672
OY	2264	GAGAGAAAGTGAACGCTGTGGAATCTCGGTGCGCGCGCGCTATGCGATGCGATGCTGTGCGACATT	2323
Db	671	GGCATGCGGTGACCTGTGGAATCTGGGCGCGGCGCTTGTGGAATGCGCTCGACGCTGC	612
OY	2324	TTGAGGGTTCGCGCGCATTCGCTCGCTAGAGTCTTTCGCGCTTTCGCTGCTGACCGCGACT	2383
Db	611	CGGCGAGCGAAGCCCA-----CTATTCGACATCTCGCGGCGGTGTGTGTTCGCGCGACT	558
OY	2384	GAATCCCGGTGGGCTGCGCTGGCGGACTTCAGGCGCATCAGGCGC---GCGGTGTGGTGA	2440
Db	557	GGGTGGCCCTTGGACTGTGCGCGCGCGCGCTGCGCGCACGTTGTGCGCAAGGCTTGGCGCTTGC	498
OY	2441	TCAGCTTGGCGCGGCGCAACCGAAGCTGCATGTGTCATCGGGTACCCCGTAGAAGC	2500
Db	497	ATGTGCTGGGTGGCGCTTACCGAAGCGGCGATCTGTGTCGAACCTGCAGAGCGTCAATACG	438
OY	2501	TCGACCTTATTCGTTGGCGCGAGATCCCTTACGCGCGCTGCGCTGCGCAACAGACGTTCCAGC	2566
Db	437	TGCGCGCGCATGTGGCTTCGATTCCTTACGCGCGCGCATTTGCGGGAACAGGCTTACCGGG	378
OY	2561	TGCTTCGATGAGCGCGCTCGAACCGCGCGCGCTTCGTGGTTTCGCGGCGCACTTACATTTGCG	2620
Db	377	TGCTGCAACCCACGAGGCGCGACGTCGCGGACCTGTGTGTGGCGGAGCTGTGGAATCGGCG	318
OY	2621	GGGTGGGCTGCACTGGGCTTACTGGCGGATGAAGAGACGCGCAAGACTTCTTCG	2680

[illegible]

```

RESULT 15
US-09-252-991A-4764
; Sequence 4764, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4764
; LENGTH: 2844
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4764

```

Query Match	8.0%;	Score 340;	DB 4;	Length 2844;
Best Local Similarity	51.9%;	Pred. No. 1.3e-65;		
Matches 917;	Conservative	0;	Mismatches 805;	Indels 44;
				Gaps 5;

Qy	1127	AGCGTATTCAAGACAGACTGTGGGAAACGATGATCACTGCGACGTAAAGCGTATCGAGG	1186
Db	1	AGACTTCCAGCGGAACCTCCACGGAGCCATTCGACACAGCCCGCATTTCCCGCCCTTGAGG	60
Qy	1187	TCCAGCGAGAGGCGCCCGGGTCTCTGGGGATCCAAAGAGCGCATTTGTTCCCGTGTGC	1246
Db	61	TGCTCCCGAGGCGCGCGG-----CAGGGGCCAGCCAGCGTCGCGCCCGGTGTGTTCG	115
Qy	1247	TCAGAGCGGCGCTTAACCGAAGTCGTGGTGTGCACTGGTTGCAGAGGCTGGGAATC	1306
Db	116	CCACCAACTGGGCGAGAGAGGCTTCGTCGCCGGGCGCTTCCGAGCGTTTCGGCAATC	175
Qy	1307	CGGTGTACACGACGACGACGACACTCTGACGTGCTGTGATCATCAGCTCTACGAGACG	1366
Db	176	TTCACGACATGCTCTCCGAGACCCCGGAGGTGTGGCTCGACACACACTCTACCGGGTGG	235
Qy	1367	ATGGGGACCTTCCTCGCGTGGGACATCGTCAGCGAAGTGTTCGCGCCGACCTTCTGG	1426
Db	236	GCGACGGTATCTCTGCTGGCGTGGGACAGCGTCGTGGCTGTGTTCCCGAAGGCGTCGCGG	295
Qy	1427	ACGCAATGCTTGAGAGCGTACGTCGTTTTTCTCCGCGCGGCTCACTGAAGAACATGGGGTG	1486
Db	296	AAACCATGTTGAAGCTTACCTAGGCGCTGTCCACGCGTCTTGCGACAGCGCTTGGAGC	355
Qy	1487	AACAGGTGCGCTGTTCGCTTCGCGCTGCCACGTAGAGAGCGCGGCGACGCAACGCGGA	1546
Db	356	AGCCCGCGATCTGCTGCTGTGCGCTTGGGCGCAGCAGCGCGCGCGGCGCTGTGTCACGCGC	415

QY 1547 CCACCGCGCTGTGACGACGATACGCTGACGCGCTGTTCGCGCGCGGCTCGAGCAC 1606
Db 416 AGCCGCGATGCGCCACGCGCGCGCACCTGCGATCGGACCTTCTTCTTCGCGCGCGGAG 475
QY 1607 TGCCCATGACGCTCGCGGTGCTGCGCGCGCGGACGACGCTGACGACGACGACGACGACG 1666
Db 476 GCGCGGATGCGGACGCGCTGCTTATCCGACCAAGTGTCAACCGCGCGGACGACGACG 535
QY 1667 GCGCTTCGCGGACGCTTCGCGCGCGCGCGCGGACGACGACGACGACGACGACGACG 1726
Db 536 AGCGTGCCTGCGGATCGCGCGCGCGCGCGGACGACGACGACGACGACGACGACG 595
QY 1727 TCGCGGTGATGAGGAGAAAGGCTGAGAGAGAGGCTTGTGCGGCTTCGCGGTCTCGAGT 1786
Db 596 TCGAGGTGAGGCTGCGCGCGCGGACGCGAGAGGTGCGCGCGGTATTCGCGGTCTCGCG 655
QY 1787 CAGGCGCGGCTTACGCGCGGATTCGATGCGCGGACGCTTCGCGGAGGCTTACGCTTCC 1846
Db 656 CAGGCGCTGCTGACGCGCGCGCTGAGCATGACGACGCGCGCGCGCGCGCGCGCTGATCG 715
QY 1847 TCGATCATGTGAGGTAAAGCTGCTGACGCGACGACGACGCTGATGCGAACTGTGAT 1906
Db 716 AAGAGGCGCGCGGCGGTATGCTGCGGATCATCGAGAGAGAGAGATCCGAGGCTTGC 775
QY 1907 GCGCGCGGAGTTCAGCGCGCTGCTGAGCGAGCGCGCGCTCGAGGCGGACGCGGAC 1966
Db 776 GCGCGCTGAGTGTCCAGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826
QY 1967 AGCTCCGATGATGCGGCTTACGACGCTTCGCGGATTCGCGGTATTCATCTTACCTCG 2026
Db 827 -----TGCCTGCGCGCGCGGACGAGGCGAGTGCCTATGTGATTCACCTCG 871
QY 2027 GATCCACAGGCTTCGCGGAGGAGGATGATGATGATGATGATGATGATGATGATGATGAT 2086
Db 872 GCTCCACGCGGCGCGGACGAGGCGCTGAGGTGAGGACGCGCGCGGATCAATACCATCG 931
QY 2087 TGGACATCAAGAGCGCTTCGAAATAGGCGCGGAGACAGAGGTGCTGCGCTTCCTCGC 2146
Db 932 AGCGCTGCTGACGCTGCTGCGGAGGAGCGCATCGGATCGCTTGTGCGGCTTCGCGC 991
QY 2147 TGAAGTTGATCTTCGCGTCTATGATGTGTTCGCGGATCTTCGCGCGCGCGGCTACGATCG 2206
Db 992 TGGATTTGATCTTCGCGTCTTTCGACCTGTTTCGCGCGCTTCGCGCGCGGCTTCG 1051
QY 2207 TGGTCCGAGCGCGCTTCAGCTGCGGATTCGCGCGGATTCGCGGATTCGAGTTCGAG 2266
Db 1052 TCTGCGCGCGGACGAGAGCGCGGATGCGCTGCTGCGGAGGCTTATCGCGCGC 1111
QY 2267 AGAAGGTGACGCGTGTGAACTTCGCGCGCGGCTGATGCGGATGCTGCTGAGCATTTTG 2326
Db 1112 ATGCGGTGAGCTGTGAACTTCGCGCGCGGCTTGTGAGATGCGCTTCAGCGCTGCGCG 1171
QY 2327 AGGCTGCGCGGATTCGCTTCGCTAGGTCTTCGCGCGCTTCGCTGCTGAGCGCGAGTGA 2386
Db 1172 CGAGCGAGCGGAC-----CTATCGAGTCTGCGGCGGAGTGTGCTGCTGCGCGAGCTG 1225
QY 2387 TCCGCGTGGCGCTGCTGCGGAGCTTCAGGCGCATGAGCGC---GCGGTGCTGAGTCA 2443
Db 1226 TGGCGCTGAGCTTCGCGCGCGCGCTGCGCGCATGTTGCGGAGGCTGCGCGCTGCA 1285
QY 2444 GCTTGGGCGGAGGACGCGAGGCTGATTCGCTTCATGCGGTACCGCTGAGGAGAGCTCG 2503
Db 1286 TGTGCGGTGCGGTACGAGAGGCGGATCTGTGCGACCTGCGAGCGGTGCGATACGCTG 1345
QY 2504 ACCTATGCTGCGGAGGATTCCTTACGCGCGCTTCGCTGCGGACGAGAGCTTCACTG 2563
Db 1346 GCGCGACTGCGGTTCGATTCCTTACGCGCGCGCATTCGCGGACGCGCTTACCGGCTG 1405
QY 2564 TCGATGAGGCGCTTCGAGCGCGCGCTGCGGTTCGCGGCGCACTTACATGCGGCGG 2623
Db 1406 TCGACACCGAGCGGCGGACGCGCGGACCTGTGCGGCGAGCTGTGATCGCGCGG 1465
QY 2624 TCGGCGTGGCACTGCGGTACTGCGCGGATGAGAGAGACGCGCAAGCTTCTCTGCTG 2683

Db 1466 CCAGCTGCGCGCGGCTTATGCAACGATCCCGAACTTCAGGCGCGCGGCTTCTG----- 1520
QY 2684 ACCCGAGACCGGAGGAGCGCTTACAAAGACGCGGATTCGCGCGGCTTACCTGCGCGATG 2743
Db 1521 -CCAGAGTCCAGGCGCGCTGCTGATGCGACCGGCGATGCGGTGCTTACGCGGCGAGC 1579
QY 2744 GAAACATGAGTTCATGCGGCGGTAGAGCAACCAATCAAGCTTTCGCGGATACCGCGTTG 2803
Db 1580 GTACCTGAGATTCCTGCGGTGCGGTGCGACGAGAGGTGAGAAAGTGGCGGCGGCGATCG 1639
QY 2804 AGCTCGGAGAAATGAGGAAAGCTCAAGTGCATTCGAACTTACGCGGCGGATTCG 2863
Db 1640 AGTTGGCGGAGGTGAGGCGCGCTGCTGCGCGCGCGGCTGCGGATGAGCGCTGCGCG 1699
QY 2864 TGCCGCTGCGGAGAGCGCGCGGAGC 2889
Db 1700 CGGTGCTGCGGCGGTGCGGTGCGGAGC 1725

Search completed: October 4, 2003, 03:40:15
Job time : 267.251 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: October 3, 2003, 11:29:26 ; Search time 1035.8 Seconds
(without alignment)
11031.749 Million cell updates/sec

Title: US-10-014-717-1_COPY_11872_16104

Perfect score: 4233
Sequence: 1 atgcagcaccatcagctctc.....aggcgagagcagctagctaa 4233

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4233	100.0	68750	21	AAZ55887
2	4114.6	97.2	71989	21	AAA29349
3	728	17.2	58857	21	AAA58471
4	543.4	12.8	4305	22	ABA89198
5	543.4	12.8	48715	22	ABA89188
6	543.4	12.8	50538	24	AB878890
7	390.4	9.2	32160	24	AB878887
8	360.6	8.5	135638	25	ABX34289

C	9	358.2	8.5	4411529	22	AA199682
C	10	349.2	8.2	4403765	22	AA199683
C	11	305.6	7.2	1416	23	AA53964
C	12	284.2	6.7	7335	22	AA51367
C	13	265	6.3	37856	21	AA11992
C	14	261.4	6.2	13029	23	AA51470
C	15	247.6	5.8	30600	24	ABQ78872
C	16	245	5.8	7110	22	AA51361
C	17	241.6	5.7	77536	21	AA14651
C	18	239	5.6	88421	24	AA140781
C	19	221.8	5.2	7347	23	AA54136
C	20	220.6	5.2	5305	22	AB89272
C	21	211.2	5.0	1566	22	AA526321
C	22	203.8	4.8	135638	25	ABX34289
C	23	202.4	4.8	5451	22	AA526319
C	24	202.2	4.8	1705	22	AB89272
C	25	188	4.4	3348	22	AB89272
C	26	181.4	4.3	1341	22	AA526322
C	27	179.4	4.2	18660	21	AA58472
C	28	177.4	4.2	4341	25	AB132151
C	29	177.4	4.2	32539	25	AB132159
C	30	172.6	4.1	7178	21	AA59145
C	31	172.2	4.1	3849	22	AA52795
C	32	171.6	4.1	23666	12	AAQ10190
C	33	168.4	4.0	8301	22	AA59035
C	34	168.4	4.0	34071	22	AA590033
C	35	168.4	4.0	42717	22	AA590032
C	36	158.2	3.7	1178	20	AA569562
C	37	155.2	3.7	3471	24	ABK74876
C	38	155	3.7	1200	20	AA569559
C	39	152.6	3.6	6390	22	AA581370
C	40	150.8	3.6	1204	20	AA569560
C	41	148.4	3.5	1803	25	AB132150
C	42	138.6	3.3	31122	14	AAQ40706
C	43	137.4	3.2	1172	20	AA569527
C	44	137.4	3.2	4615	22	AA590034
C	45	136.6	3.2	3798	24	ABK74884

ALIGNMENTS

AAZ55887	standard; DNA; 68750 BP.
AAZ55887	
10-APR-2000	(first entry)
Sorangium cellulosum 68.75 kb contig.	
Epochlone biosynthesis; type I polyketide synthase; taxol substructure; anticancer; ds.	
Sorangium cellulosum.	
Key	Location/Qualifiers
CDS	1..1826
FT	/*tag= a
FT	/partial
FT	/product= "partial Orf 1 protein (AA558580)"
FT	/note= "No initiation codon given in the specification"
FT	complement (1900..3171)
FT	/*tag= b
FT	/product= "Orf 2 protein (AA558581)"
FT	3415..5556
FT	/*tag= c
FT	/product= "Orf 3 protein (AA558582)"
FT	complement (5612..5992)
FT	/*tag= d
FT	/product= "Orf 4 protein (AA558583)"
FT	6226..6675
FT	CDS

Mycobacterium tube
Mycobacterium tube
Klebsiella pneumoniae
Quorum sensing con
S. cellulosum DNA
Pseudomonas aeruginosa
S. roseosporus dep
Quorum sensing con
Nucleotide sequenc
88421nt genomic DN
Pseudomonas aeruginosa
Escherichia coli p
Pseudomonas sp 11p
Nucleotide sequenc
Benzodiazepines bl
Benzodiazepines bl
DNA encoding a pep
S. chrysomallus ac
Cephalosporin anti
Nucleotide sequenc
Nucleotide sequenc
Benzodiazepines bl
Benzodiazepines bl
Soil derived pepti
Bacillus lichenif
Soil derived pepti
Quorum sensing con
Soil derived pepti
Benzodiazepines bl
Bacillus subtilis
Soil bacteria pept
Nucleotide sequenc
Bacillus lichenif

```

FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      7610..11875
CDS      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
CDS      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
CDS      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
CDS      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
CDS      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
CDS      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
CDS      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
CDS      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
CDS      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
CDS      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
CDS      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
CDS      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
CDS      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
CDS      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
CDS      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
CDS      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
CDS      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX
XX      WO9966028-A2.
XX
XX      23-DEC-1999.
XX
XX      16-JUN-1999; 99WO-EP04171.
XX
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
XX      05-FEB-1999; 99US-0118906.

```

```

PA      (NOVS ) NOVARTIS AG.
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERN GES MBH.
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX      WPI: 2000-097741/08.
DR      P-PDB: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
DR      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
DR      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
DR      AAV58592, AAV58593, AAV58594.
XX
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy
XX
XX      Claim 14, Page 87-104; 174pp; English.
XX
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-58578) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiplex drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
XX
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX      Query Match 100.0%; Score 4233; DB 21; Length 68750;
XX      Best Local Similarity 100.0%; Pred. No. 0;
XX      Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX      1 ATGACGATCAATCACTCTGTAACAGCTCGACGACGAGGATCACTGGCGCGAT 60
XX      11872 ATGACGATCAATCACTCTGTAACAGCTCGACGACGAGGATCACTGGCGCGAT 11931
XX
XX      61 GGGGAGCGCTTCAGATACAGGCGCCCAAGAAAGCGCTTAAACCGGACCTGCTCGA 120
XX      11932 GGGGAGCGCTTCAGATACAGGCGCCCAAGAAAGCGCTTAAACCGGACCTGCTCGA 11991
XX
XX      121 ATCTCGAGCACAAAGACAGATCTGACGATGCTCTGTCAGAGATCTCCGCAATCC 180
XX      11992 ATCTCGAGCACAAAGACAGATCTGACGATGCTCTGTCAGAGATCTCCGCAATCC 12051
XX
XX      181 ATCTCGAGCGCGCGCGCGCGAGGCGGAGCGGCTCGTCTCTCAACAGATCAAGATCC 240
XX      12052 ATCTCGAGCGCGCGCGCGAGGCGGAGCGGCTCGTCTCTCAACAGATCAAGATCC 12111
XX
XX      241 TACTGCTGGCGCGGACAGAGCGCTTACGATCGTCCAGCGGAGATCCAGCGCTATCGGAA 300
XX      12112 TACTGCTGGCGCGGACAGAGCGCTTACGATCGTCCAGCGGAGATCCAGCGCTATCGGAA 12171
XX
XX      301 TAGACTGTAGGATCTTGAAGTGGCGGAGCTGAGCGCGCGCTTTCGGAAGATCGTCCGG 360
XX      12172 TAGACTGTAGGATCTTGAAGTGGCGGAGCTGAGCGCGCGCTTTCGGAAGATCGTCCGG 12231
XX      361 CGGACGACATGCTTGGGCGCCACACAGCGTGGCCGACATGATCAGTGATCAGAGCTAAA 420

```


Db 12232 CGGACGACATGCTTCGGGCCCCACACGCTGCCGACATGATCCAGGTATCGAGCTTAA 12291
Qy 421 GTCCGCGCGACATCGATGATCATCATGCGCGGGGCTCGACCGGAGACAACGGGAACG 480
Db 12292 GTCCAGCGCGACATCGATCATCATGCGCGGGGCTCGACCGGAGACAACGGGAACG 12351
Qy 481 AGGCTCGTGTGCTTGCAGATGCGATGTGCGACCGCATCTATGACACCGAGCGCCCTCG 540
Db 12352 AGGCTCGTGTGCTTGCAGATGCGATGTGCGACCGCATCTATGACACCGAGCGCCCTCG 12411
Qy 541 CTCTATCATGCTGTGCGCGCTTGCCTGCGACGAGCGGCAAAACCGTCTGTGTCTCATGTC 600
Db 12412 CTCTATCATGCTGTGCGCGCTTGCCTGCGACGAGCGGCAAAACCGTCTGTGTCTCATGTC 12471
Qy 601 GATCTCATTTAAAGTTGACCTTAGGCGAGCCGTGCATCATCTTCAAGACCTGGCTCAGTTC 660
Db 12472 GATCTCATTTAAAGTTGACCTTAGGCGAGCCGTGCATCATCTTCAAGACCTGGCTCAGTTC 12531
Qy 661 TACGAAAGATCCCGAGACCTCTCTCCCTGTCTGAGAGCTCTCGTACCGGCAATTATGATCTC 720
Db 12532 TACGAAAGATCCCGAGACCTCTCTCCCTGTCTGAGAGCTCTCGTACCGGCAATTATGATCTC 12591
Qy 721 GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATCAACGATCGATGATTAATTGAGCGG 780
Db 12592 GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATCAACGATCGATGATTAATTGAGCGG 12651
Qy 781 CGCANTCGCGAGCTCCGACCTCCCGCGAGAGCTTCGAGTGAAGGCGGACATCTACCTCCG 840
Db 12652 CGCANTCGCGAGCTCCGACCTCCCGCGAGAGCTTCGAGTGAAGGCGGACATCTACCTCCG 12711
Qy 841 AAGGAGATCCGCTTCGCGACAACGAGAGATGCTGCGCTCGGACTCTCGGAGGCTGATG 900
Db 12712 AAGGAGATCCGCTTCGCGACAACGAGAGATGCTGCGCTCGGACTCTCGGAGGCTGATG 12771
Qy 901 AAGCGCGTGTGCGGAGACCGCGGCTGAACCCCGACGCGGCTGATCTGAGCTGATTTCC 960
Db 12772 AAGCGCGTGTGCGGAGACCGCGGCTGAACCCCGACGCGGCTGATCTGAGCTGATTTCC 12831
Qy 961 GAGGATCGGCGCGTGAAGCGCGGAGCCCGGTTTAAAGCTCAATAACCTCTTCAAC 1020
Db 12832 GAGGATCGGCGCGTGAAGCGCGGAGCCCGGTTTAAAGCTCAATAACCTCTTCAAC 12891
Qy 1021 CGGCTCCCCGTCATCCCGCGCTGAACGATATCAACGGGAGCTTCACTGATGCTCTC 1080
Db 12892 CGGCTCCCCGTCATCCCGCGCTGAACGATATCAACGGGAGCTTCACTGATGCTCTC 12951
Qy 1081 CTGACATCGACACCACTCGCGACAAGAGCTTTCGAACAGCGGCTTAAGCTTCAAGAG 1140
Db 12952 CTGACATCGACACCACTCGCGACAAGAGCTTTCGAACAGCGGCTTAAGCTTCAAGAG 13011
Qy 1141 CAGCTGTGGGAAGCGATGATATCATGCGACGTAACGCGTATGAGGTCCACGAGAGCC 1200
Db 13012 CAGCTGTGGGAAGCGATGATATCATGCGACGTAACGCGTATGAGGTCCACGAGAGCC 13071
Qy 1201 GCCCGGTCCTGAGGATATCAACGAGGCGCATTTGTTCCCGTGTGCTCAACGAGCGCTT 1260
Db 13072 GCCCGGTCCTGAGGATATCAACGAGGCGCATTTGTTCCCGTGTGCTCAACGAGCGCTT 13131
Qy 1261 AACGCAAGTCTGTTGATGTCACCTGTTGACGAGGCTCGAATCCGCTGTACACAGC 1320
Db 13132 AACGCAAGTCTGTTGATGTCACCTGTTGACGAGGCTCGAATCCGCTGTACACAGC 13191
Qy 1321 ACGGAGATCTCTCAGCTGCTGTGATATCATGACGCTTACGAGACAGATGGGAGCTGCTC 1380
Db 13192 ACGGAGATCTCTCAGCTGCTGTGATATCATGACGCTTACGAGACAGATGGGAGCTGCTC 13251
Qy 1381 CTCGCTGGGACATGCTGACGAGATGTTCCCGCGCACTTCTGAGACGACATGCTGAA 1440
Db 13252 CTCGCTGGGACATGCTGACGAGATGTTCCCGCGCACTTCTGAGACGACATGCTGAA 13311
Qy 1441 GCGTACGTGTTTTCTCCGCGGCTCATCTGAGAACATGGGATGAAACAGGTGCGCTGT 1500

Db 13312 GCGTACGTGTTTTCTCCGCGGCTCATGAGAACCATGGGTGAACAGTGGCTGT 13371
Qy 1501 TCGCTTCGCTGCGCCAGACTGAGAGCGGGGCGAGCGCAACCGACCGCTGCTG 1560
Db 13372 TCGCTTCGCTGCGCCAGACTGAGAGCGGGGCGAGCGCAACCGACCGCTGCTG 13431
Qy 1561 AGGAGCATACGCTGCAACGCGCTGTTCGCGCGCGGCTGAGCAGCTGCCCATGACGTC 1620
Db 13432 AGGAGCATACGCTGCAACGCGCGCTGTTCGCGCGCGGCTGAGCAGCTGCCCATGACGTC 13491
Qy 1621 GCCGTGTGTGCGCGCGCAAGACGCTCAACGAAAGCTTTGGCGCCGTTCCGGGGA 1680
Db 13492 GCCGTGTGTGCGCGCGCAAGACGCTCAACGAAAGCTTTGGCGCCGTTCCGGGGA 13551
Qy 1681 CTTGAGCGCGCGCTGCGCGAGACGAGGAGCGCCGGAACAATTGCTGCGGTGATG 1740
Db 13552 CTTGAGCGCGCGCTGCGCGAGACGAGGAGCGCCGGAACAATTGCTGCGGTGATG 13611
Qy 1741 GAGAAAGCTGGAGACAGTTGTGCGCGGTTCTCGCGTGTCTGAGTCAGGCGCGCTTAC 1800
Db 13612 GAGAAAGCTGGAGACAGTTGTGCGCGGTTCTCGCGTGTCTGAGTCAGGCGCGCTTAC 13671
Qy 1801 GTGCGGATCGATGCGCGACCTTACCGGCGGAGGCTATCCACTCTCTGATCATGCTGAG 1860
Db 13672 GTGCGGATCGATGCGCGACCTTACCGGCGGAGGCTATCCACTCTCTGATCATGCTGAG 13731
Qy 1861 GTAAAGCTCGTGTGACGACGACCTAGGCTGATGAGCAAACTGTATGAGCGCGCGGATC 1920
Db 13732 GTAAAGCTCGTGTGACGACGACCTAGGCTGATGAGCAAACTGTATGAGCGCGCGGATC 13791
Qy 1921 CAGCGGCTGCTCTGAGACGAGGCGCGGCTCGAAGGCGACGCGGACCACTCTCGATGATG 1980
Db 13792 CAGCGGCTGCTCTGAGACGAGGCGCGGCTCGAAGGCGACGCGGACCACTCTCGATGATG 13851
Qy 1381 CCCATTGACACCTTGGATCTTGCGGATGTCTATCTTACACTGCGGATCCACAGGCTTG 2040
Db 13852 CCCATTGACACCTTGGATCTTGCGGATGTCTATCTTACACTGCGGATCCACAGGCTTG 13911
Qy 2041 CCCAAGGGGAGTATGATCGATCATCGGGGATGCGTCAACACATCTGGAATCAACAGAG 2100
Db 13912 CCCAAGGGGAGTATGATCGATCATCGGGGATGCGTCAACACATCTGGAATCAACAGAG 13971
Qy 2101 CGCTTCAAAATAGGAGCCCGGAGACAGGAGTGTGCGCTCTCTCGCTGAGCTTGAATCTC 2160
Db 13972 CGCTTCAAAATAGGAGCCCGGAGACAGGAGTGTGCGCTCTCTCGCTGAGCTTGAATCTC 14031
Qy 2161 TCGGCTATATGATGTTGCTGGGATCTTGCGCGCGGCGGTACGATCTGTGTGCTCGAGCGG 2220
Db 14032 TCGGCTATATGATGTTGCTGGGATCTTGCGCGCGGCGGTACGATCTGTGTGCTCGAGCGG 14091
Qy 2221 TCCAAAGCTGCGGATCCGGGCGCATTTGGGACAGATTGATCGAACGAGAGAGTGCAGG 2280
Db 14092 TCCAAAGCTGCGGATCCGGGCGCATTTGGGACAGATTGATCGAACGAGAGAGTGCAGG 14151
Qy 2281 TGAACCTGAGTCCGGGCGCTGATGCGATGCTCTGACGCAATTTTGAAGGCTCGGCCGAT 2340
Db 14152 TGAACCTGAGTCCGGGCGCTGATGCGATGCTCTGACGCAATTTTGAAGGCTCGGCCGAT 14211
Qy 2341 TCGCTCGTATGCTCTGCGGCTTTTCTGCTGTGACGCGGCACTGATCCCGTGGGCTTG 2400
Db 14212 TCGCTCGTATGCTCTGCGGCTTTTCTGCTGTGACGCGGCACTGATCCCGTGGGCTTG 14271
Qy 2401 CTTGCGAGCTCCAGGCGCATTAAGGCGCGGCGTGTGCGTATCAACCTGGGGGGGCGCAC 2460
Db 14272 CTTGCGAGCTCCAGGCGCATTAAGGCGCGGCGTGTGCGTATCAACCTGGGGGGGCGCAC 14331
Qy 2461 GAAAGCTGATCTGTGCTCATGAGGTAACCCCTGAGGAACGTCGATCTGAGGCGAGC 2520
Db 14332 GAAAGCTGATCTGTGCTCATGAGGTAACCCCTGAGGAACGTCGATCTGAGGCGAGC 14391
Qy 2521 ATCCCTTACCGGCGCTGCTGCGCAACGACGTTTCCAGTGTCTGATGAGGCGCTGAA 2580
Db 14392 ATCCCTTACCGGCGCTGCTGCGCAACGACGTTTCCAGTGTCTGATGAGGCGCTGAA 14451

QY	2581	CCGCGCCCGGCTGGGTTCCGGGGCAACTCAATTTGGCGGGTCCGGCTGGCACTGGCC	2640
Db	14482	CCGCGCCCGGCTGGGTTCCGGGGCAACTCAATTTGGCGGGTCCGGCTGGCACTGGCC	14511
QY	2641	TACTGGCCGATGAAGAGAAACGCGAGAGCTTCTCGTGCACCCCGAGACCGGGGAG	2700
Db	14512	TACTGGCCGATGAAGAGAAACCGGAGAGCTTCTCGTGCACCCCGAGACCGGGGAG	14571
QY	2701	CGCCTTCAAGAACCGGGGATCTTGGGCGCTACCTCGCCGATGGAAACATCGAGTTCAAG	2760
Db	14572	CGCCTTCAAGAACCGGGGATCTTGGGCGCTACCTCGCCGATGGAAACATCGAGTTCAAG	14631
QY	2761	GGGGGTGAGCAACCAATCAACTTCGCGGATACCGGCTTGAGCTCGGGGAAATCGAG	2820
Db	14632	GGGGGTGAGCAACCAATCAACTTCGCGGATACCGGCTTGAGCTCGGGGAAATCGAG	14691
QY	2821	GAATCGCTCAAGTGGCATCCGACGTAACGCGACGCGGTGATTGTGCCCTCGGAAACGAC	2880
Db	14692	GAATCGCTCAAGTGGCATCCGACGTAACGCGACGCGGTGATTGTGCCCTCGGAAACGAC	14751
QY	2881	GGGGGGAACAAGCTCCCTTCTAGCCTTAATGTGGTCCCGGAGGGACAAGGAAACGCGTGGC	2940
Db	14752	GGGGGGAACAAGCTCCCTTCTAGCCTTAATGTGGTCCCGGAGGGACAAGGAAACGCGTGGC	14811
QY	2941	GAGCAGACCGGAGCTCTCAAGACCGGACGGATCGACGCGAGACACACGCGCCGGAAGCG	3000
Db	14812	GAGCAGACCGGAGCTCTCAAGACCGGACGGATCGACGCGAGACACACGCGCCGGAAGCG	14871
QY	3001	GACGGCTTGAACGACGCGGAGAGGGTCACTTCAAGTCCGCTGACACGGAATCCGGAAG	3060
Db	14872	GACGGCTTGAACGACGCGGAGAGGGTCACTTCAAGTCCGCTGACACGGAATCCGGAAG	14931
QY	3061	GACCTGACGAGAAAGCCGTCGTGATCTGACCGGGACGAGATCCGCGGAGCGGGGCTG	3120
Db	14932	GACCTGACGAGAAAGCCGTCGTGATCTGACCGGGACGAGATCCGCGGAGCGGGGCTG	14991
QY	3121	GACGCTTACGCGCGCTCGCCGTAAGCTTCTTGAAGCCCGGATTCGGTTGTT	3180
Db	14992	GACGCTTACGCGCGCGCTCGCCGTAAGCTTCTTGAAGCCCGGATTCGGTTGTT	15051
QY	3181	GAGTTTGGTCGATTCTGAGCTGCTTGAAGCGGTGAGGCCGACGGCGGACCCCTTCCC	3240
Db	15052	GAGTTTGGTCGATTCTGAGCTGCTTGAAGCGGTGAGGCCGACGGCGGACCCCTTCCC	15111
QY	3241	AAATTCGGTTATCCATCGGCGGGACGACGTAACCCGGTGCMAACTACGCGTATGTCAAA	3300
Db	15112	AAATTCGGTTATCCATCGGCGGGACGACGTAACCCGGTGCMAACTACGCGTATGTCAAA	15171
QY	3301	TCCGGCCGCATTCGAGGGGTGGAAGAGGCTTCAATTAACAACCCGTTTCAGGCAACCT	3360
Db	15172	TCCGGCCGCATTCGAGGGGTGGAAGAGGCTTCAATTAACAACCCGTTTCAGGCAACCT	15231
QY	3361	TTTGCTGAAGCTCTCCGATCAGGGATGAGCGCGGACGCACTTGCGGCAAAACTTCGAC	3420
Db	15232	TTTGCTGAAGCTCTCCGATCAGGGATGAGCGCGGACGCACTTGCGGCAAAACTTCGAC	15291
QY	3421	GTGTTCCGATGAAGCGGCGTTCAACCTCTGTTGTTGGGCGAGATCGACGCCATCGAGTGC	3480
Db	15292	GTGTTCCGATGAAGCGGCGTTCAACCTCTGTTGTTGGGCGAGATCGACGCCATCGAGTGC	15351
QY	3481	CTGTATGGAATGTCGTGCGCGAGATTTTGGCTGCGTGAAGCGGATTTATGGCGAGCTTC	3540
Db	15352	CTGTATGGAATGTCGTGCGCGAGATTTTGGCTGCGTGAAGCGGATTTATGGCGAGCTTC	15411
QY	3541	CTGTATGAGACGAGCGCCTTCTCTGCACATCGCGCTCTGTCCGTCGGGCGCAATTCATTTT	3600
Db	15412	CTGTATGAGACGAGCGCCTTCTCTGCACATCGCGCTCTGTCCGTCGGGCGCAATTCATTTT	15471
QY	3601	GAACAGGTTCCGGCGGTTCTCGACCTCGACATTCGAGCGTTTACGTCACCGCATGCTG	3660
Db	15472	GAACAGGTTCCGGCGGTTCTCGACCTCGACATTCGAGCGTTTACGTCACCGCATGCTG	15531

Qy	3661	GGGCGGGGGGTAAACCCGCGGCGAGTTCCAGAGTCTGAACGCTCGGTACGAGATTCTTCACCG	3720
Db	15532	GGCGGGGGGGTAAACCCGCGGCGAGTTCCAGAGTCTGAACGCTCGGTACGAGATTCTTCACCG	15591
Qy	3721	AGGCGCGCCCAACGACGCGCGCGCCCTCCCGCGCGAGCAGACCTTCGCGGATATGCTT	3780
Db	15592	AGGCGCGCCCAACGACGCGCGCGCCCTCCCGCGCGAGCAGACCTTCGCGGATATGCTT	15651
Qy	3781	CGGACATTTCTTGAAGAACAACTAACCCGAGTACATGTGTCTTACATCTTTCTGTGAAGCTC	3840
Db	15652	CGGACATTTCTTGAAGAACAACTAACCCGAGTACATGTGTCTTACATCTTTCTGTGAAGCTC	15711
Qy	3841	GATGCGTTTGCCTGACGTCTCAACGGGCAAGTGTGATGCTTAAGGCCCTGCGCGAGCGAAG	3900
Db	15712	GATGCGTTTGCCTGACGTCTCAACGGGCAAGTGTGATGCTTAAGGCCCTGCGCGAGCGAAG	15771
Qy	3901	GATACCTCTGTGCGCGCGGCACTTTCGGGAGCACACGCGCCCAACGGGACGCTTGGAGAGATC	3960
Db	15772	GATACCTCTGTGCGCGCGGCACTTTCGGGAGCACACGCGCCCAACGGGACGCTTGGAGAGATC	15831
Qy	3961	CTCGTCGCGGTCTGTACGGGAGGTGCTTCGGGCTTGGAGTGTGTGGGCTCCAGACAGCTTC	4020
Db	15832	CTCGTCGCGGTCTGTACGGGAGGTGCTTCGGGCTTGGAGTGTGTGGGCTCCAGACAGCTTC	15891
Qy	4021	GTCGATCTTGTGTCGACATCGATTCACATCTTCGATAGAGAGCCTGTGTGAGAGAGAG	4080
Db	15892	GTCGATCTTGTGTCGACATCGATTCACATCTTCGATAGAGAGCCTGTGTGAGAGAGAG	15951
Qy	4081	CTGAGATAGGAGATGCCCATCAACCGAGTTGTTCCAGTACCCGAACTCGGCTCGCTGAGCG	4140
Db	15952	CTGAGATAGGAGATGCCCATCAACCGAGTTGTTCCAGTACCCGAACTCGGCTCGCTGAGCG	16011
Qy	4141	TCCGATTTCGCGCGGAGACTTCGAGAGATCTAAGATCAGCGGCCGAAACATGACAGACCGAGTG	4200
Db	16012	TCCGATTTCGCGCGGAGACTTCGAGAGATCTAAGATCAGCGGCCGAAACATGACAGACCGAGTG	16071
Qy	4201	GAGGTTTCGCGCAAGGCGACGAGACCTTACTTAA	4233
Db	16072	GAGGTTTCGCGCAAGGCGACGAGACCTTACTTAA	16104
RESULT 2			
AAA29349			
ID	AAA29349	standard; DNA; 71989 BP.	
XX	XX		
AC	AAA29349;		
XX	XX		
DT	12-SEP-2000	(first entry)	
XX	XX		
DE	Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.		
XX	XX		
KW	Epothilone; polyketide synthase; epOA; epOB; epOC; epOD; epOE; epOF; epOL; epOK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal; tubulin polymerization assay; anti-tumour; cytostatic; de.		
KW	XX		
OS	Sorangium cellulosum.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	CDS	3..992	
FT		/+tag= a	
FT		/label= ORF_A	
FT		/product= transposase	
FT		/note= "not part of the PKS"	
FT	CDS	989..1501	
FT		/+tag= b	
FT		/label= ORF_B	
FT		/product= transposase	
FT		/note= "not part of the PKS"	
FT	CDS	1998..6263	
FT		/+tag= c	
FT		/label= epOA_gene	
FT		/note= "encodes the loading domain"	
FT	misc RNA	2031..3548	

```
FT      /*tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      misc_RNA
FT      3621..4661
FT      /*tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      misc_RNA
FT      4917..5810
FT      /*tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      misc_RNA
FT      5856..6155
FT      /*tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      CDS
FT      6260..10493
FT      /*tag= h
FT      /label= epob_gene
FT      /note= "encodes module 1, the NRPS module"
FT      misc_RNA
FT      2031..3548
FT      /*tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      2031..3548
FT      /*tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      misc_RNA
FT      6861..6887
FT      /*tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      misc_RNA
FT      6861..6887
FT      /*tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      misc_RNA
FT      7358..7366
FT      /*tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      7898..7921
FT      /*tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      misc_RNA
FT      8261..8308
FT      /*tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      misc_RNA
FT      8411..8422
FT      /*tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      misc_RNA
FT      8861..8905
FT      /*tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      misc_RNA
FT      8966..8983
FT      /*tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      misc_RNA
FT      9090..9179
FT      /*tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      misc_RNA
FT      9183..9992
FT      /*tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      misc_RNA
FT      10121..10138
FT      /*tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      misc_RNA
FT      10261..10306
FT      /*tag= w
FT      /note= "encodes thiolation domain (PCP) of the NRPS
FT      module"
FT      CDS
FT      10639..16137
FT      /*tag= x
```

```
FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      misc_RNA
FT      10654..12033
FT      /*tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      misc_RNA
FT      12250..13287
FT      /*tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      misc_RNA
FT      13327..13899
FT      /*tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      misc_RNA
FT      14962..15756
FT      /*tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      misc_RNA
FT      15763..16008
FT      /*tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      CDS
FT      16134..37907
FT      /*tag= ad
FT      /label= epod_gene
FT      /note= "encodes modules 3-6"
FT      misc_RNA
FT      16425..17606
FT      /*tag= ae
FT      /note= "encodes KS3"
FT      misc_RNA
FT      17817..18857
FT      /*tag= af
FT      /note= "encodes AT3"
FT      misc_RNA
FT      19581..20396
FT      /*tag= ag
FT      /note= "encodes KR3"
FT      misc_RNA
FT      20424..20642
FT      /*tag= ah
FT      /note= "encodes ACP3"
FT      misc_RNA
FT      20706..22082
FT      /*tag= ai
FT      /note= "encodes KS4"
FT      misc_RNA
FT      22296..23336
FT      /*tag= aj
FT      /note= "encodes AT4"
FT      misc_RNA
FT      24069..24647
FT      /*tag= ak
FT      /note= "encodes KR4"
FT      misc_RNA
FT      24867..25151
FT      /*tag= al
FT      /note= "encodes ACP4"
FT      misc_RNA
FT      25203..26576
FT      /*tag= am
FT      /note= "encodes KS5"
FT      misc_RNA
FT      26793..27883
FT      /*tag= an
FT      /note= "encodes AT5"
FT      misc_RNA
FT      27966..28574
FT      /*tag= ao
FT      /note= "encodes DHS"
FT      misc_RNA
FT      29433..30287
FT      /*tag= ap
FT      /note= "encodes ER5"
FT      misc_RNA
FT      30321..30869
FT      /*tag= aq
FT      /note= "encodes KR5"
FT      misc_RNA
FT      31077..31373
FT      /*tag= ar
FT      /note= "encodes ACP5"
FT      misc_RNA
FT      31440..32807
FT      /*tag= as
FT      /note= "encodes KS6"
FT      misc_RNA
FT      33018..34067
FT      /*tag= at
FT      /note= "encodes AT6"
FT      misc_RNA
FT      34107..34676
FT      /*tag= au
```

FT /note="encodes DH6"
FT misc_RNA 35760..35641
FT /tag= av
FT /note="encodes ER6"
FT misc_RNA 36705..37256
FT /tag= aw
FT /note="encodes KR6"
FT misc_RNA 37470..37769
FT /tag= ax
FT /note="encodes ACP6"
FT CDS 37912..49308
FT /tag= ay
FT /label= epOR gene
FT /note="encodes modules 7 and 8"
FT misc_RNA 38014..39375
FT /tag= az
FT /note="encodes KS7"
FT misc_RNA 39589..40626
FT /tag= ba
FT /note="encodes AT7"
FT misc_RNA 41341..41922
FT /tag= db
FT /note="encodes KR7"
FT misc_RNA 42181..42423

Query Match 97.2%; Score 4114.6; DB 21; Length 71989;
Best Local Similarity 98.3%; Pired. No. 0;

Matches 4159; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 ATGACATCAATCAGCTTCTGAACGAGCTCGACACAGGATATCAAGCTGGGGCGGAT 60
DB 6260 ATGACATCAATCAGCTTCTGAACGAGCTCGACACAGGATATCAAGCTGGGGCGGAT 6319
QY 61 GGGGAGCGCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCGAACCCTGCTCGTCA 120
DB 6320 GGGGAGCGCTCCAGATACAGGCCCCCAAGAACGCCCTGAACCGAACCCTGCTCGTCA 6379
QY 121 ATCTCCGAGACCAAAAGCAGATCTCTGAGATGCTCCGTGAGAGATCTCCCGGAGATCC 180
DB 6380 ATCTCCGAGACCAAAAGCAGATCTCTGAGATGCTCCGTGAGAGATCTCCCGGAGATCC 6439
QY 181 ATCGTCCCGCCCGCAGCGGAGCGGAGCTCGTTCCTCTCAGACATCAAGAAATCC 240
DB 6440 ATCGTCCCGCCCGCAGCGGAGCGGAGCTCGTTCCTCTCAGACATCAAGAAATCC 6499
QY 241 TACTGCTGGGCGGAGCAGAGGCTTTACGCTCCGAGCGGATCAACGCTATCGGAA 300
DB 6500 TACTGCTGGGCGGAGCAGAGGCTTTACGCTCCGAGCGGATCAACGCTATCGGAA 6559
QY 301 TACGACTGTAACGATATCTGACGTCGCGAGCTGAGCGCGCTTTGGGAAAGTCGTCCG 360
DB 6560 TACGACTGTAACGATATCTGACGTCGCGAGCTGAGCGCGCTTTGGGAAAGTCGTCCG 6619
QY 361 CGGACGACATGCTTGGGGCCACACGCTGCCGACATGATGAGGTGATCGAGCTTAA 420
DB 6620 CGGACGACATGCTTGGGGCCACACGCTGCCGACATGATGAGGTGATCGAGCTTAA 6679
QY 421 GTGAGCGCGGACATCGAGATCATGATCTGCGGGGCTGACCGGAGCAACGCGAAAGCG 480
DB 6680 GTGAGCGCGGACATCGAGATCATGATCTGCGGGGCTGACCGGAGCAACGCGAAAGCG 6739
QY 481 AGGCTGCTGCTTGGGAGATGCGATGTCGACCCGATCTATGACACCGAGCCCTCCG 540
DB 6740 AGGCTGCTGCTTGGGAGATGCGATGTCGACCCGATCTATGACACCGAGCCCTCCG 6799
QY 541 CTCTATCAGCTGCTGCGGCTTGGGCTGAGAGCGGAGAACCCGCTCTGCTCAGTATC 600
DB 6800 CTCTATCAGCTGCTGCGGCTTGGGCTGAGAGCGGAGAACCCGCTCTGCTCAGTATC 6859
QY 601 GATCTCATTAACCTTGAACCTAGGCAAGCTGCTCATCATCTTCAAGAGCTGGCTCAGCTTC 660
DB 6860 GATCTCATTAACCTTGAACCTAGGCAAGCTGCTCATCATCTTCAAGAGCTGGCTCAGCTTC 6919

QY 661 TACGAAGATCCGAGACCTCTCCCTGTCGAGCTCTGTACCGGATTAATGATCTC 720
DB 6920 TACGAAGATCCGAGACCTCTCTCCCTGTCGAGCTCTGTACCGGATTAATGATCTC 6979
QY 721 GCGCTGAGCTCTGCAAGAGATCTGAGGCGCATCAACGATGATGATTTACTGAAAGCGG 780
DB 6980 GCGCTGAGCTCTGCAAGAGATCTGAGGCGCATCAACGATGATGATTTACTGAAAGCGG 7039
QY 781 CGCATGCGGAGCTCCACCTCCCGGAGCGCTTCGATGAAAGCCGATCTCATCTACCTG 840
DB 7040 CGCATGCGGAGCTCCACCTCCCGGAGCGCTTCGATGAAAGCCGATCTCATCTACCTG 7099
QY 841 AAGGAGATCCGCTTCGAGACACGAGACATGCTGCGTGGAGCTCTGAGGATGATG 900
DB 7100 AAGGAGATCCGCTTCGAGACACGAGACATGCTGCGTGGAGCTCTGAGGATGATG 7159
QY 901 AAGCGCGCTGTCGAGGAGCGCGGAGCTGACCCCGAGCGGCGTATCTTGGCTGATTTTC 960
DB 7160 AAGCGCGCTGTCGAGGAGCGCGGAGCTGACCCCGAGCGGCGTATCTTGGCTGATTTTC 7219
QY 961 GAGGTGATGAGGCGCTGAGGCGCGGAGCGGCGGCTTGAAGCTTGAAGCTTGAAG 1020
DB 7220 GAGGTGATGAGGCGCTGAGGCGCGGAGCGGCGGCTTGAAGCTTGAAGCTTGAAG 7279
QY 1021 CGGCTCCCGCTCATCCGCGGCTGAACGATATCACCGGAGCTTCAAGCTGATGCTTC 1080
DB 7280 CGGCTCCCGCTCATCCGCGGCTGAACGATATCACCGGAGCTTCAAGCTGATGCTTC 7339
QY 1081 CTGACATCGACACCACTCGCGCAAGAGCTTGAAGCGGCTTGAAGCTTGAAG 1140
DB 7340 CTGACATCGACACCACTCGCGCAAGAGCTTGAAGCGGCTTGAAGCTTGAAG 7399
QY 1141 CAGCTGTGGAAGCGATGATATCTGCGACGTAAGCGGATGAGCTCAGCGAGAGGCT 1200
DB 7400 CAGCTGTGGAAGCGATGATATCTGCGACGTAAGCGGATGAGCTCAGCGAGAGGCT 7459
QY 1201 GCGCGGCTCTGAGGAGATCCAAAGAGCGGATTTCCCGTGGTCTCAGAGCGGCTT 1260
DB 7460 GCGCGGCTCTGAGGAGATCCAAAGAGCGGATTTCCCGTGGTCTCAGAGCGGCTT 7519
QY 1261 AACGACAAAGTGTGTGTCACTCTGTTGACAGAGCTCGAATCCGCTGTACACGAC 1320
DB 7520 AACGACAAAGTGTGTGTGTCACTCTGTTGACAGAGCTCGAATCCGCTGTACACGAC 7579
QY 1321 AGCGACATCTCTGACGCTGTCTGATCATGAGCTCTTACAGACAGATGAGGAGCTCTGCT 1380
DB 7580 AGCGACATCTCTGACGCTGTCTGATCATGAGCTCTTACAGACAGATGAGGAGCTCTGCT 7639
QY 1381 CTGCGGTGGAACATGCTGACGAGGTGTTCCGCGGACCTTCTGAGACGATGCTGAA 1440
DB 7640 CTGCGGTGGAACATGCTGACGAGGTGTTCCGCGGACCTTCTGAGACGATGCTGAA 7699
QY 1441 GCGTACGTGTTTTCTCCGAGGCTCACTGAGAAACATGAGGAGTGAACAGGTGCGCTGT 1500
DB 7700 GCGTACGTGTTTTCTCCGAGGCTCACTGAGAAACATGAGGAGTGAACAGGTGCGCTGT 7759
QY 1501 TGGCTTCCGCTCCGCAAGTAAAGCGGAGGAGCGCAAAAGCGACCAACCGCTGCTG 1560
DB 7760 TGGCTTCCGCTCCGCAAGTAAAGCGGAGGAGCGCAAAAGCGACCAACCGCTGCTG 7819
QY 1561 AAGGACATACGCTGACGAGCTGTTCCGCGGAGGAGTGAACAGCTGCCATGACGCTC 1620
DB 7820 AAGGACATACGCTGACGAGCTGTTCCGCGGAGGAGTGAACAGCTGCCATGACGCTC 7879
QY 1621 GCGTGTGTGCGCGCGCAAGACGCTCAGTACGAAGCTTTCCGCGCTTCCGCGCGA 1680
DB 7880 GCGTGTGTGCGCGCGCAAGACGCTCAGTACGAAGCTTTCCGCGCTTCCGCGCGA 7939
QY 1681 CTTGGCGCGGCTGCGAGAGGAGGAGCGCGCGCAACACTTGTGCGGCTGAGTGA 1740
DB 7940 CTTGGCGCGGCTGCGAGAGGAGGAGCGCGCGCAACACTTGTGCGGCTGAGTGA 7999
QY 1741 GAGAAAGCTGGGAGAGGATGTGCGGCTTCTCGCGTGTCTGAGTCAAGCGCGGCTTAC 1800

|||||
Db 8000 GAGAAAGCTGGAGACAGGTTGTCGGGCTTCGGGGTGTGAGTCAAGCGCGGCTAC 8059
OY 1801 GGGCGGATCGATCGGACCTACCGGCGGAGGGATCCACTACTCTCTGATATGTTAG 1860
Db 8060 GTGGCGATCGATCGGACCTACCGGCGGAGGGATCCACTACTCTCTGATATGTTAG 8119
OY 1861 GTAAAGCTCTGTCGATCGGACCATGATGAGCAAACTGATCACTGAGCGCGGAGATC 1920
Db 8120 GTAAAGCTCTGTCGATCGGACCATGATGAGCAAACTGATCACTGAGCGCGGAGATC 8179
OY 1921 CAGCGGCTGCTGTGAGCGAGCGCGGCTGAAAGCGAGCGGACCGAGCTTCGATGATG 1980
Db 8180 CAGCGGCTGCTGTGAGCGAGCGCGGCTGAAAGCGAGCGGACCGAGCTTCGATGATG 8239
OY 1981 CCCATTCAAGACCTTCGGATCTCGGATATGTCATCTACCTCGGATCCACAGGTTG 2040
Db 8240 CCCATTCAAGACCTTCGGATCTCGGATATGTCATCTACCTCGGATCCACAGGTTG 8299
OY 2041 CCCAAGGGGGTATGATCGATCGGAGTGGCGCTCAACACCATCTCGATCAACAG 2100
Db 8300 CCCAAGGGGGTATGATCGATCGGAGTGGCGCTCAACACCATCTCGATCAACAG 8359
OY 2101 CGCTTCGAAATAGGCGCCGAGACAGAGTGTGCGCGCTCTCTCGCTGAGCTTCATCTC 2160
Db 8360 CGCTTCGAAATAGGCGCCGAGACAGAGTGTGCGCGCTCTCTCGCTGAGCTTCATCTC 8419
OY 2161 TCGGCTATGATGATGTCGGGATCTCGGCGGCGGCGGATAGATGTCGTCGGAACCG 2220
Db 8420 TCGGCTATGATGATGTCGGGATCTCGGCGGCGGCGGATAGATGTCGTCGGAACCG 8479
OY 2221 TCCAAGCTGCGGATCGGCGGATTTGGGAGAGTTGATGAAAGAGAGAGTGAACG 2280
Db 8480 TCCAAGCTGCGGATCGGCGGATTTGGGAGAGTTGATGAAAGAGAGAGTGAACG 8539
OY 2281 TGGAACTGCGTCCGCGCTGATCGGATGCTGTCGAGCATTTTGAAGGTCGCCCAT 2340
Db 8540 TGGAACTGCGTCCGCGCTGATCGGATGCTGTCGAGCATTTTGAAGGTCGCCCAT 8599
OY 2341 TCGGCTGATGATGTCGGGATCTCGGCGGCGGCGGATAGATGTCGTCGGAACCG 2400
Db 8600 TCGGCTGATGATGTCGGGATCTCGGCGGCGGCGGATAGATGTCGTCGGAACCG 8659
OY 2401 CTGCGGAGCTCCAGGCTCAGGCGCGGCTGTGCTGATGATGAGCTGCGGCGGCGCAC 2460
Db 8660 CTGCGGAGCTCCAGGCTCAGGCGCGGCTGTGCTGATGATGAGCTGCGGCGGCGCAC 8719
OY 2461 GAAAGCTGATGTCGATCGGATCCCGGATAGAGGATGATGATGTCGTCGTCGAGC 2520
Db 8720 GAAAGCTGATGTCGATCGGATCCCGGATAGAGGATGATGATGTCGTCGTCGAGC 8779
OY 2521 ATCCCTACGCGGCTCCGTCGCGCAACAGAGTTCCAGTCTGATGAGCGCTCGA 2580
Db 8780 ATCCCTACGCGGCTCCGTCGCGCAACAGAGTTCCAGTCTGATGAGCGCTCGA 8839
OY 2581 CCGCGCGGATGAGGATCCGCGGCAACTCTCATTTGGGCGGATGCGGCTGCGACCTG 2640
Db 8840 CCGCGCGGATGAGGATCCGCGGCAACTCTCATTTGGGCGGATGCGGCTGCGACCTG 8899
OY 2641 TACTGCGGATGAG 2700
Db 8900 TACTGCGGATGAG 8959
OY 2701 CGGCTCTACAGACCGGCGATCTGGGCGGCTACCTGCGGATGAGAACTAGATTCATG 2760
Db 8960 CGGCTCTACAGACCGGCGATCTGGGCGGCTACCTGCGGATGAGAACTAGATTCATG 9019
OY 2761 GGGCGTGAAGCAACCAATCAAGCTTCGCGATACCGGCTGAGCTCGGAGAAATCGAG 2820
Db 9020 GGGCGTGAAGCAACCAATCAAGCTTCGCGATACCGGCTGAGCTCGGAGAAATCGAG 9079
OY 2821 GAAAGCTCAAGTCCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
|||||

Db 9080 GAAAGCTCAAGTCGATCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9139
OY 2881 GGGCGCAACAGCTCTCTGAGCTATGTCGCGAGAGAGAGAGAGAGAGAGAGAGAG 2940
Db 9140 GGGCGCAACAGCTCTCTGAGCTATGTCGCGAGAGAGAGAGAGAGAGAGAGAGAG 9199
OY 2941 GAGCAG 3000
Db 9200 GAGCAG 9259
OY 3001 GAGCGCTTGAAG 3060
Db 9260 GAGCGATTAAG 9319
OY 3061 GAGCTGAAG 3120
Db 9320 GATCTGAG 9379
OY 3121 GAGCTGAAG 3180
Db 9380 GAGCTGAAG 9439
OY 3181 GAGCTTGAAG 3240
Db 9440 GAGCTTGAAG 9499
OY 3241 GAGCTTGAAG 3300
Db 9500 GAGCTTGAAG 9559
OY 3301 TCCGCGCGATCGAG 3360
Db 9560 TCCGCGCGATCGAG 9619
OY 3361 TTGCTGAAG 3420
Db 9620 TTGCTGAAG 9679
OY 3421 GTGCTGAAG 3480
Db 9680 GTGCTGAAG 9739
OY 3481 CTGATGAG 3540
Db 9740 CTGATGAG 9799
OY 3541 CTGATGAG 3600
Db 9800 CTGATGAG 9859
OY 3601 GAAAGCTTGGGCGGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Db 9860 GAAAGCTTGGGCGGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9919
OY 3661 GGGCGGAG 3720
Db 9920 GGGCGGAG 9979
OY 3721 AGGCGGAG 3780
Db 9980 AGGCGGAG 10039
OY 3781 CGGCACTTCTGAG 3840
Db 10040 CGGCACTTCTGAG 10099
OY 3841 GATGCTTGGGCGGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 10100 GATGCTTGGGCGGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10159
OY 3901 GATGCTTGGGCGGAG 3960
Db 10160 GATGCTTGGGCGGAG 10219
|||||

Qy	3961	CTCGTCGGCGATGCTACCGGAGAGTGTCTGGGCTGGAGGTGGTCTCGGGCTCCAGACGACTTC	4020
Db	10220	CTCGTTGGCGATGCTACCGGAGAGTGTCTGGGCTGGAGGTGGTCTCGGGCTCCAGACGACTTC	10278
Qy	4021	GTCGATCTTGGTGGCAGATCGCATTCACATCGTTCGATGAGAGCCGTGTCAGAAAGAG	4080
Db	10280	GTCGATCTTGGTGGCAGATCGCATTCACATCGTTCGATGAGAGGTGTGTCAGAAAGAG	10339
Qy	4081	CTGGAATGAGGAGATCGCATTCACCGAGTGTGTTCCAGTACCCGAACTCGCTGCTGGGG	4140
Db	10340	CTGGAATGAGGAGATCGCATTCACCGAGTGTGTTCCAGTACCCGAACTCGCTGCTGGGG	10399
Qy	4141	TCCGATTGGCGCCGAGACTCGAGAGATCTAGATAGAGGCGCGAACAATGACGACCGAGTG	4200
Db	10400	TCCGATTGGCGCCGAGACTCGAAGATCTTAGAGCAGGCGCCGCAATGACGACCGAGTG	10455
Qy	4201	GAGGTTCCGCGCAAGGCGAGAGAGACTGACTAA	4233
Db	10460	GAGGTTCCGCGCGAAGGCGAGAGAGACTGACTAA	10492
RESULT 3			
AAAS8471	standard; DNA; 58857 BP.		
AAAS8471			
AAAS8471;			
31-OCT-2000	(first entry)		
Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.			
BLM gene cluster; bleomycin gene cluster; polyketide metabolite; bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine; thiazoline; dithiazoline; microbial metabolite; sugar; ss.			
Streptomyces verticillius.			
Key	location/Qualifiers		
223..564			
/*tag= a	/transl_except= (pos: 1..3, aa: Met)		
/*tag= b	/transl_except= (pos: 1..3, aa: Met)		
/*tag= c	/transl_except= (pos: 1..3, aa: Met)		
/*tag= d	/transl_except= (pos: 1..3, aa: Met)		
/*tag= e	/transl_except= (pos: 1..3, aa: Met)		
/*tag= f	/transl_except= (pos: 1..3, aa: Met)		
/*tag= g	/transl_except= (pos: 1..3, aa: Met)		
/*tag= h	/transl_except= (pos: 1..3, aa: Met)		
/*tag= i	/transl_except= (pos: 1..3, aa: Met)		
/*tag= j	/transl_except= (pos: 1..3, aa: Met)		
/*tag= k	/transl_except= (pos: 1..3, aa: Met)		
/*tag= l	/transl_except= (pos: 1..3, aa: Met)		
/*tag= m	/transl_except= (pos: 1..3, aa: Met)		
/*tag= n	/transl_except= (pos: 1..3, aa: Met)		
/*tag= o	/transl_except= (pos: 1..3, aa: Met)		
/*tag= p	/transl_except= (pos: 1..3, aa: Met)		
/*tag= q	/transl_except= (pos: 1..3, aa: Met)		
/*tag= r	/transl_except= (pos: 1..3, aa: Met)		
/*tag= s	/transl_except= (pos: 1..3, aa: Met)		
/*tag= t	/transl_except= (pos: 1..3, aa: Met)		
/*tag= u	/transl_except= (pos: 1..3, aa: Met)		
/*tag= v	/transl_except= (pos: 1..3, aa: Met)		
/*tag= w	/transl_except= (pos: 1..3, aa: Met)		
/*tag= x	/transl_except= (pos: 1..3, aa: Met)		
/*tag= y	/transl_except= (pos: 1..3, aa: Met)		
/*tag= z	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 0	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 1	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 2	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 3	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 4	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 5	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 6	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 7	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 8	/transl_except= (pos: 1..3, aa: Met)		
/*tag= 9	/transl_except= (pos: 1..3, aa: Met)		
/*tag= A	/transl_except= (pos: 1..3, aa: Met)		
/*tag= B	/transl_except= (pos: 1..3, aa: Met)		
/*tag= C	/transl_except= (pos: 1..3, aa: Met)		
/*tag= D	/transl_except= (pos: 1..3, aa: Met)		
/*tag= E	/transl_except= (pos: 1..3, aa: Met)		
/*tag= F	/transl_except= (pos: 1..3, aa: Met)		
/*tag= G	/transl_except= (pos: 1..3, aa: Met)		
/*tag= H	/transl_except= (pos: 1..3, aa: Met)		
/*tag= I	/transl_except= (pos: 1..3, aa: Met)		
/*tag= J	/transl_except= (pos: 1..3, aa: Met)		
/*tag= K			

```

FT CDS 34827..35804
FT /*tag= x
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 20; encodes AAB07566"
FT CDS 35818..37302
FT /*tag= 1
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 19; encodes AAB07567"
FT CDS 37299..39215
FT /*tag= m
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 18; encodes AAB07568"
FT CDS 39301..47181
FT /*tag= n
FT /note= "ORF 17; encodes AAB07569"
FT CDS 47178..49985
FT /*tag= o
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 16; encodes AAB07570"
FT CDS 49982..51001
FT /*tag= p
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 15; encodes AAB07571"
FT CDS 50998..52386
FT /*tag= q
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 14; encodes AAB07572"
FT CDS 52383..52946
FT /*tag= r
FT /note= "ORF 13; encodes AAB07573"
FT CDS 53018..54190
FT /*tag= B
FT /note= "ORF 12; encodes AAB07574"
FT CDS 54187..55824
FT /*tag= t
FT /note= "ORF 11; encodes AAB07575"
FT CDS 55821..56093
FT /*tag= u
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 10; encodes AAB07576"
FT CDS 56090..57586
FT /*tag= v
FT /transl_except= (pos: 1..3, aa: Met)
FT /note= "ORF 9; encodes AAB07577"
FT CDS 57583..58857
FT /*tag= w
FT /note= "ORF 8; encodes AAB07578"

MO2000040704-A1.
XX PN 13-JUL-2000.
XX PD
XX
XX 06-JAN-2000; 2000MO-US00445.
XX PF
XX 06-JAN-1999; 99US-0115435.
XX PR 05-FEB-1999; 99US-0118848.
XX PR 05-JAN-2000; 2000US-0477962.
XX
XX (REGC ) UNIV CALIFORNIA.
XX PA
XX SHEN B, DU L, SANCHEZ C, CHEN M, EDWARDS DJ;
XX
XX WPI: 2000-465974/40.
XX DR P-PSDB: AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
XX DR AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568
XX DR AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
XX DR AAB07576, AAB07577, AAB07578.
XX
XX New bleomycin gene cluster components useful for peptide and/or
XX PT polypeptide metabolites, especially bleomycin, production and for
XX PT chemically modifying biological molecules -
XX
XX Claim 8; Page 97-136; 163pp; English.
XX

```


Db 44297 TCTTCACTCGGCTCCACCGGAAACCAAGGCGCTCATGATGACACACTTCGGGCGG 44356
 QY 2075 TCACACCATCTGACATCAACGAGCGCTTCGAAATAGGCCCGGAGACAGGCTGCTG 2134
 Db 44357 CCACACACCTGGAAATCGTCAACCGCGCTTCGGACCGGCGCTTCGGCTGCTG 44416
 QY 2135 CGCTCCCTCGGCTGACCTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2194
 Db 44417 CGCTCTCTCTCCGACCTTCGACCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44476
 QY 2195 GCGGTACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2254
 Db 44477 GCGGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44536
 QY 2255 TGATGCAAG 2314
 Db 44537 TGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 44596
 QY 2315 TCGAGCATTTTGAAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2374
 Db 44597 CCGAGTACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 44656
 QY 2375 GCGGCGACTGATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2431
 Db 44657 GCGGCGACTGATCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44716
 QY 2432 TGTGCGTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2491
 Db 44717 CCACCGTATGAGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 44776
 QY 2492 TGAAGAACCTGCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2551
 Db 44777 TCGGGAAGGTGCAAGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 44836
 QY 2552 CGTTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2611
 Db 44837 GGTGGAAGGTCTCTGAG 44896
 QY 2612 ACATTGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2671
 Db 44897 ACATTGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44956
 QY 2672 GCTTCCTGCTGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2731
 Db 44957 GCTTCCTGCTGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 45016
 QY 2732 ACCTGCCGATGGAACATCGAGTTCATGAGGCGGCTGAGAGACAAATCAAGCTTCGCG 2791
 Db 45017 ACCTGCCGATGGAACATCGAGTTCATGAGGCGGCTGAGAGACAAATCAAGCTTCGCG 45076
 QY 2792 GATACCGCGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2851
 Db 45077 GATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 45136
 QY 2852 AGCGGCTGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2911
 Db 45137 CCGGCGGCTGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 45196
 QY 2912 TCCCGG 2917
 Db 45197 TACCGG 45202

RESULT 4
 ABA89198
 ID ABA89198 standard; DNA; 4305 BP.
 AC
 AC ABA89198;
 XX
 XX
 DT 11-FEB-2002 (first entry)
 XX
 DE Escherichia coli polynucleotide SEQ ID NO 940.
 XX

KW Escherichia coli; B2/D+/-; anti-inflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicaemia;
 KW pyelonephritis; antibiotic resistance; ds.
 OS Escherichia coli.
 PN WO200166572-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 12-MAR-2001; 2001WO-BP03445.
 XX
 PR 10-MAR-2000; 2000FR-0003145.
 XX
 PR 02-FEB-2001; 2001FR-0001449.
 XX
 PA (INRM) INSERM INSJ NAT SANTE & RECH MEDICALE.
 XX
 PI Bingen E, Bonacorsi S, Clermont O, Nasif X, Tinsley C;
 DR WPI; 2001-550253/61.
 XX
 PT A library of DNA fragments of Escherichia coli strains for the
 PT phylogenetic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A- -
 XX
 PS Example 6; Fig 6; 646pp; English.
 XX
 CC The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
 CC B2/D+/-A-. The polynucleotides have potential anti-inflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenetic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicaemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.
 XX
 SQ Sequence 4305 BP; 880 A; 1148 C; 1309 G; 968 T; 0 other;
 Query Match 12.8%; Score 543.4; DB 22; Length 4305;
 Best Local Similarity 48.5%; Pred. No. 4,9e-106;
 Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;
 QY 181 ATCGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
 Db 241 ATAGTGCACGACCTGAGCATGTCTATCAACGCTTCCATTGACGAGATGTTACGGCGCC 300
 QY 241 TACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 Db 241 ATAGTGCACGACCTGAGCATGTCTATCAACGCTTCCATTGACGAGATGTTACGGCGCC 300
 QY 241 TACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
 Db 301 TACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 357
 QY 301 TACGATGACGAGATTCGACGCTGCCGAGGCTGACGCGGCGGCTTCGGAATTCGTCGG 360
 Db 358 TTTGACGTAGAACACTTTAATGTTACGGGTTTACCTGCGGTAAATGCGGTAGCTGCT 417
 QY 361 CCGACGACATGCTTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
 Db 418 CCGCATGAATCTACGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
 QY 421 GTTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 477
 Db 478 GTGCGCGGATACGATGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 537
 QY 478 GCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537
 Db 538 GATGCTTATGATCGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
 QY 538 CCGCTTATACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 597

Db 538 CGGCTGTTGATTTAGTTATTCGGCTTGCACGGGCGCAACATGCGCTGATTTTCAGT 657
Qy 538 ATGCATCTCATTAAGCTTGAAGCTTGAAGCGCTGTCCATCATCTTGAAGAGCTGGCGAC 657
Db 658 CTGCATCTGCTGATTTGCGCATGCTCTGATATTCGCGACCTTACAGAGAGTTGATGATG 717
Qy 658 TTCTACGAAGATCCCGAAGCTCTCTCTGCTGTCTGAGCTCTGTACCGCATTTATGTA 717
Db 718 CTGTAACCGTGAAGCCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 777
Qy 718 CTGCGGCTGAGATCTGCGAAGATCTGAGGCGCATCAACATTCATGATTTACTGGAAG 777
Db 778 CAGGGGCTGTTGTTAGAGAGCGAGTGAAGCTTATGCAACGCGATCAGGCTTATGGCA 837
Qy 778 CGGCGCATCGCGAGCTCCACCTCGCGGACGCTTCCGATGAGAGCGCATTCATCTACC 837
Db 838 CGGGGCGCTGCCGAGCTGTATGAGCCACACAGCTGCCGTACAGGCGCATTTGGCGCA 897
Qy 838 CTGAAGAGATCCGCTTCCGGGCAACGAGCAATGAGCTGCGCTGAGCTCTGAGGAGTGA 897
Db 898 CTGCTCGGATCAGGTTGATGCTGCGCGCTCATCGGCTGTGAGCCCAACAGTGGGAGTG 957
Qy 898 TTGAAGCGGCTGTGCGGGAGCGCGGCTGACCCCGAGCGGCGTATCTGCTGCTGATTT 957
Db 958 CTGACCGGCTGGCCCAACGCAACGTAATCAACAGAGCGCATTTGTTGACAGTCTTT 1017
Qy 958 TCCGAGGATGAGGGGCTGAGCGCGGAGCGCGGCTTACGCTCAACATACGCTCTTC 1017
Db 1018 AGCCAAAGTGTGACAGTGTGAGCCTTACCCGAGCTTACGCTCAATGTAGCTGTTTC 1077
Qy 1018 AACCGGCTCCCGCTCATCCGCGGCTGAGACGATATCACCGGAGACTTCACTGATGTC 1077
Db 1078 AACCGCGCGAGGTTACCCCAACGAGAGAGTATTTGATTTTACCGCTGTACG 1137
Qy 1078 CTCTGAGATGACACCACTGCGCAAGAGCTTGAACAGCGGCTTAAAGCTTCA 1137
Db 1138 TTGCTGAATGTTTGTACGACAGCAGCAGCTTATGCCCCAACAGCTGAGGTATTCAG 1197
Qy 1138 GAGCAGCTGTGGGAGAGCATGATGATCTGCGACGATGAGGCGAT-----CGAGTC 1188
Db 1198 GTGCACTGTGGGAGAGATTCGAACTGCTCTTCAAGGAGTCCGCGCCAGCAAGGG 1257
Qy 1198 CAGCAGAGCGCGCGCTGCTGAGATCCAAAGAGAGCATTTGCTCCCGTGTGCTC 1248
Db 1258 CTGATCCATAGCGGTGCTTTCATGCGCGAGTCCGGTGTATTCATAGTATGTTGGAT 1317
Qy 1249 ACGAGCGGCTTAAACAGCAAGTCTGTGTGATCACTGCTTGAAGAGCTCGGAATCTCG 1308
Db 1318 ATCGACGGGAGACAGACTGCGCAAGACCTCGGAGACAAACCGTTTACTCTGTGTCG 1377
Qy 1309 GTGTACACAGACGAGCATCTGAGCTGCTGAGATCATCACTCTACAGACAGAT 1368
Db 1378 GAGCGCAATATTATCCCAACACCGAGGTGCTGATCACAGATGATTCAGTTGGT 1437
Qy 1369 GGGGACCTGCTGCTGCGGTGAGACATCTGCGAGAGTGTCCCGCCGACCTTGTGAC 1428
Db 1438 GGGGAGTTGCAATTTCACTGGGACCGCGTCAAGCACTTTGATTAACACGCTGTGAT 1497
Qy 1429 GACATGCTGAAAGCTAGCTGTTTTTCTCCGGGCTCACTGAGAAACATAGGGGTGA 1488
Db 1498 CAGATGTTGTGCTTATGTCATGCGTGCAGGCGCTGTTGCCATGCG-----G 1548
Qy 1489 CAGGTGCGCTGTTCCTTCCGCTGCCAGCTAGAAAGCGGGCGAGCGCAACCGCAC 1548
Db 1549 CAAGTGTGAGGGGATAAATGATTTCTGCGCGCTGCCACCGTTAGTGCACGATTCAG 1608
Qy 1549 AACGCGCTGCTGAGGAGCATACGCTGACGCGCTGTTGCGGGCGGGCTGAGAGCTG 1608
Db 1609 CAGGCTCTGACACAGCGCTTGTGTTGACATGATTACTGCTGACGAGCACTGACG 1668
Qy 1609 CCCATGACGCTGCGGTGTGTGGGCGCAAGACGCTCACTGAGAAAGCTTTTCGCGC 1668

Db 1669 CCACAGAACTGCGCTGATCATGCTTATCCGTGAATTCATTCGCAACTGTGACG 1728
Qy 1669 GATTGCGGGGACTTGTGCGCGCTGCGGAGAGAGGGGACAGCGCCGAACATTTGTC 1728
Db 1729 GCGGCGATCATGTGCGCGCGCGCTTGTAGCGCTGTGGGCTGTACAGTCCGCGT 1788
Qy 1729 GCGGTGTGATGAGAAAGCTGTGGAGAGGTTGTGCGGTTCTCGCGTCTGATGTA 1788
Db 1789 GCGGTGTGATGAGAAAGCTGTGGAGAGGTTGTGCGGTTCTCGCGTCTGATGTA 1848
Qy 1789 GCGGCGCTTACGTGCGCATGATGCGCACTTACCGGGAGCGGTATTCATCTCTC 1848
Db 1849 GATGAGTGTATCTGCAAGTGAATTCGATGCTGCTGACGAGCATGCTGATGCTGATG 1908
Qy 1849 GATCATGTGAGATTAAGCTGCTGCTGACGAGCATGCTGATGCTGATGCTGATG 1908
Db 1909 ACGGTGGCGAGGTGCGGCTTCAAGTAAACAGCGCGGTCTCACCAATTTGAGCGCTG 1968
Qy 1909 CCGCGGGGATTCAGCGGCTGCTGTGAGCGAGCGCGGTGAAAGCGCGACGACG 1968
Db 1969 CTGCGCG-----TCTGATCATTCAGCAGCGGAATGTGGAACGCGCTGCTCG 2016
Qy 1969 CCTCGATGATGCCATTCAGACACTTTCGAGATCTGCGGTATGTCATTCATCTACCTCGGA 2028
Db 2017 CCGTTGCTGAAAGTGTGCGGAGTGTCAAGATGTGCGCTTATTCATTTTCACTTCGCGC 2076
Qy 2029 TCCACAGGTTGCCCAAGGGGATGATGATCATCTCGGGTCCGCTCAACACCATCTCG 2088
Db 2077 TCCACGCTTACCCGAAAGATGATGATTCACACGCTGCGGCTATTAACGCTGGA 2136
Qy 2089 GATATCAACGAGCGCTTGAATAGGGCCGGAAGACAGGTTGTGCGCTCTCTGCTG 2148
Db 2137 GATATCAACGAGCGCTTGAATAGGGCCGGAAGATGATTCACACGCTGCGGCTATTAACGCTGGA 2196
Qy 2149 AGCTTGAATCTCTGCTTATGATGTTGTTGCGGATCTGCGCGCGGCTGATGATGTCG 2208
Db 2197 AGCTTGAATCTCTGCTTATGATGTTGTTGCGGATCTGCGCGCGGCTGATGATGTCG 2256
Qy 2209 GTGCGGAGCGGCTCAAGCTGCGGATCCGCGCATTTGGGAGAGTGTGAGAAAGAG 2268
Db 2257 CTGCGGAGAGAGAGCGGAAAGATCCGCTCATTTGGCAGACATTTAGCACAGCT 2316
Qy 2269 AAGGTAGGTTGTGAACTCTGCTGCGCGCTGATGCGGATCTGCTGAGCAATTTGAG 2328
Db 2317 CATGTAACTGTGAAATGAGTGCAGCTGCGCAGTGAATGAGTGTGCG---AATACAC 2373
Qy 2329 GGTGCGCCGATGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2388
Db 2374 AGCGCGATCGGATGATGATTCGACAGTGTGCTGTGCGACATGTTGAGCGGCGATC 2433
Qy 2389 CCGGTGCGCTGCTGCGG---GCTCCAGGCACTACAGCGCGCGGTGCTGCTGATGACG 2445
Db 2434 CCGCTTAACTTACCGGAGACATGCGAGCGGCTCAATGAAGATGATGATCATCTAGT 2493
Qy 2446 CTGCGCGGCGCACCGAGCTGATCTGCTCACTGCGGTACCCCGTGAAGAACTGAC 2505
Db 2494 CTGCGTGAAGAGACGAGTGGCCATCTGTGCGGTCTTACCCGATGAGTGGTGA 2553
Qy 2506 CTATGTGAGGAGATCTCTTACGCGCGCTGCTGCGCAACAGACCTTCCAGTGTCTC 2565
Db 2554 TCGAGTGTGACAGATTTCTTACGCTGCGGCGCTGCGCAACAGCCGATATTCGCTG 2613
Qy 2566 GATGAGGCGCTGCAACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2625
Db 2614 AATGTGCACTGAGAGAAATGCTCGGTGCGGCTGAGAGAGATTTGATTTGCGGAGTG 2673
Qy 2626 GGGTGTGCACTGCGCTTACGCGGATGAGAGAAAGCGCAAGAGCTTCTGCTGAC 2685
Db 2674 GGGGTGTGCAAGGCTTACGAGAGCAAGAAAGCGGCGGAGGCTTGTGCTGCGCGC 2733
Qy 2686 CCCGAGACCGGAGAGCGCTTACAGAGCGCGGATCTGCGGCTGCTGCTGCGGATGGA 2745
Db 2734 GAGGAGTGTGTGAGCGAATTTACCGCACTGCGGATTCGCGGCGCTACTTGTGACGCG 2793

QY	2746	AACATCGAGTTTCATGGGCGGTGAGGACAACCAAAATCAAGCTTCGGGATATCCGCGTTAG	2805
Db	2794	CAAGTCGCGCTTTTGGGGCGCAGCATACCAAGTAAGGTGAAATGGTTTACCGTATCGAA	2853
QY	2806	CTCGGGGAAATCGAGAAAACGTCAATCGATCCGAACGTACCGGACCGGATGATGTG	2865
Db	2854	CTGGGGGAAATCGAGCGCGCTGCATTTCCGGACATCCCGATGTGGAGCACTGATGGTGGTG	2913
QY	2866	CCCGTCGGGAAACGACGGCGGCAACAAGCTCTTTCAGCTATGTGTGTCCCGAAGGGCA	2925
Db	2914	GCAGGGGTAA-----TTCTCAATCTGTCCG	2940
QY	2926	CGGAGACCGCGCTCGGACGAGACCGAGCCTCAGACCGAGCGGATCGACCGCAGACA	2985
Db	2941	CTGTGTCGCTTTTGGCAAACTGCAGCATGGCCACCGGCGGAGGCATTGCCAAGTAAAGAA	3000
QY	2986	CACCGCCCGAAGCGGACGGCTTGAGCGACCGCGAGAGGTGACGTTCAAGCTCGCTCGA	3045
Db	3001	-----GGGAGGCGCGCGGCACTGGCGGAGGGTATTAATGTGAAATCCGGCAGAGCTCTA	3054
QY	3046	CACGGACTCCGGAGGGACCTGGACGGAAAGCCCGCTGTCAATCTACCGGGGAGGATCCG	3105
Db	3055	GGCTTCAAACTCAAGAGACCACTATTTGGCGGCTGTGATGTGCTTGGGCAATTGCATGACG	3114
QY	3106	CGGAGGCGGGGCGTGAAGCTTACGCGCGGTGCGGTAGCGTTCGAAAGTTCCTTGAAGCC	3165
Db	3115	GCACCGGGGATAGACACAGTTTACATCAAAAGCCGCGACATATCTGATTTAGCGCGCA	3174
QY	3166	CCGATTCGGTTTGTGATGTGTGTGTCGATTTCTGAGCTCTTGAGGAGGCTGAGCCGAC	3225
Db	3175	AAACACACGCTGGCACAGTGGGGAAATGCTGTGCGGCTTGGGCAAGTGGTTACCC	3234
QY	3226	GGCGGCAAGCCCTTCCCAAAATTCGCTATCCATCGGCGGCGACGACTACCCGATGCAAC	3285
Db	3235	GGTCACTTTTGGCAAGTATGCTATGGTCCGCGGGGGCTATACCCGCTGCAAC	3294
QY	3286	TACCGCTATGTCAAAATCCGGCCGCATCGAGGCGTGAACGAGGCTTCTATTTATTCAC	3345
Db	3295	TACGCTACCTGCACTTCAGACACAGATTCGAAAGGGAGATCCGGAATTTACTACTTCCAC	3354
QY	3346	CCGTTCGAGCAACCGTTTGTGTAAGCT--CTCCGATCAACGGGATGAGCGGAGCCGAC	3402
Db	3355	CCGGACAGAGCTGTCTTATGCCGTTAGCACCAAGATCGAGCTAAACATGGTTTAT	3414
QY	3403	GTTCCGCAAACTTCGACGCTGTTCGATGAAGCGCGTTCAACTTCCTGTTCGTGGCAGG	3462
Db	3415	GCCGGACCTTAATCACTATTATGCGCATCGGGCGGATTCACGCTGTTATGTGTGCTGAT	3474
QY	3463	ATCGACGCCATCGAATCGCTGTGATG---ATGCTGTGGCCAGAAATTTTGTCTGTGAG	3519
Db	3475	ATGGGGGATCTCCCAATTCATGTGGCAAGGAGAGCTTGGCACTTCTCGGTATGGAA	3534
QY	3520	GCCGGATATATGSCCAGCTCTCTGATGAGCAGCGGCTTCTCTCGAACAATCGGCGCTGT	3579
Db	3535	GCAGGTACTCTCTGCAATTTACTGGAAGAAATGGCGCGCTACGATTTGGGGCTGTG	3594
QY	3580	CCGGTGGGCAATTAATTTTGAACAGTTTCGCGCGGTTCTTCGACTTCGACATTTCCGAC	3639
Db	3595	CAACTTGGGATGGCAGACTTTTTCGCTGTGGCATGCAATTTTCATTTGTCCGACATCAT	3654
QY	3640	GTTTATGTGACCGGATGCTGGGCGGGGAGTGAACCGCGGCACTTCCAGGTCTGTACG	3699
Db	3655	CGCTATGTCTCATTCGACCGTGGGGGGCGCGATGGGCAAGAGCGGCAAGTCTGACGA	3714
QY	3700	CTCGGTCAAGATTCTTCACCGAGCGCGCACGACGCGGCGGCCCTCCGCGCGCAG	3755
Db	3715	TTGCTGGCGGATTTCTCCACCTATGAGAAACGGAAGAAACCGCTGCGCGCTGGACATG	3774
QY	3760	CAGCACTTCGCGGATATGTCTTGGCACTTTCTTGAAGACCAAACTACCGGATGATGTGTG	3819
Db	3775	CAGACTCAAAAGATGCCATGTGTGGCGGCGCTGCTCAGCAAACTGCTGCACTATATGTGTG	3834

Oy	3820	CCCTACAGTCTTCGTGAGAGCTGATGCGTGGCCGTGACGTCCAAACGGCAAGTGCATCGT	3879
Dd	3835	CCGATGATCTGATGTTATAGCCACCGATTTTCCGTAAACCGTAAACGGCAAGCTGATGCG	3894
Oy	3880	AAGGCGCTGCGCAGCGAGGAGATACCTCGTGGCGCGGCAT-----TCGGGCAACAG	3933
Dd	3895	CAAAATTTACAGCTGCAAGGGGGAACAAATTGCCACAGCGGTGACGGCGTGGTCCAAATC	3954
Oy	3934	GCGCCACGGGACGCGCTTGGAGAGATCTCGTCGCGGTCTGACGGAGAGTCTCGGCGTG	3993
Dd	3955	CAGGTGACAGTGGCTTAAACAACAGCGCGTGGTGGCGCTCTGGCAGAGAGTACTCGGCGTG	4014
Oy	3994	GAGGTGGTCGGGCTTCACAGACGCTTGTGCATCTTGTGTCGACATCGATTCAATCGTT	4053
Dd	4015	AGCCACGCTTCGGCCGAAGACGATTTCTTCGCTGGGGGGCGACGTTCTATGAAATTGGTGTG	4074
Oy	4054	CGCATGAGAGACCTCTTTCAGAAAGAGCGTGATAGGAGATCGGCATTCACCGAGTTGTTT	4113
Dd	4075	CGTATTCAGCAGGCACTTGAGGCGCATTAATCGGCGAGAGATTCCTCATTTGTGCATCTGTTC	4134
Oy	4114	CAGTACCGAAGCTCGGCTCGCTGGCG	4140
Dd	4135	CGTCTGCCAACCATTCGCGGATGTGGCG	4161

RESULT 5

ID ABA89188 standard; DNA; 48715 BP

AC ABA89188;

DT 11-FEB-2002 (first entry)

DE Escherichia coli polynucleotide SEQ ID NO 921.

KW Escherichia coli, B2/D+A-; antiinflammatory; antibacterial;

KW systemic infection; non-diarrhoeal infection; septicaemia;

XX

XX

XX

XX

XX

PR 02-FEB-2001; 2001FR-0001449.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE

PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

DR WPI; 2001-550253/61.

PT A library of DNA fragments of *Escherichia coli* strains for the

PT nature B2/D+ A- -

PS Example 6; Fig 6; 646pp; English

CC The invention relates to a library of DNA fragments of *Escherichia coli*

CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature

CC antibacterial and immunosuppressive activity as part of pharmaceutical

CC infections. The polypeptides are useful for determining the phylogenetic

CC an undesired development of *E. coli*, particularly an extra-intestinal

CC Septicemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.

XX Sequence 48715 BP; 11934 A; 13488 C; 12203 G; 11090 T; 0 other;

Query Match 12.8%; Score 543.4; DB 22; Length 48715;
Best Local Similarity 48.5%; Pred. No. 8.5e-106;
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;

```
OY 181 ATCCGCCCCCGCCCGAGGCGACGCTCCGTTCTCTCTACAGACATCCAGAAATCC 240
DB 18373 ATAGGACAGACCCCTGAGCATGTCATCAACCGTTGCCATTGACGAAATGTCAGGCCCC 18314
OY 241 TACTGGCTGGGCGGACAGAGCGTTTACGGTCCCGAGGGATCCAGCCCTATCGCGAA 300
DB 18313 TACTGGCTTGGAGC---ACAGACGGGTGCTACTGATCGGACCCACATCTTACCATGAA 18257
OY 301 TACGACTGTAGCGATCTCGACGTGCGAGGCTGAGCCGCGCCTTTCGAAAGTCTGCG 360
DB 18256 TTTGACGTAGAACATTTAATGTTACGGGTTTACCATGCGGGTAATGCGCTGATCCGT 18197
OY 361 CGGACGACATGCTCTGGGCGCCACACGCTGCCCGACATGATCAGTATGAGCTTAA 420
DB 18196 CGCATGAAATGCTACGTCGCGGGTACTCCCGACGGTACTACAGACAAATCTTGCGCA 18137
OY 421 GTCCAGCG---CGACATCGAAGATCATGATCTGCGCGGGCTGACCGGAGCACAGGGA 477
DB 18136 GTGCGGCGGTACATTAAGAGACGCGCATCTGATGCTTGTCTCCCTTACGACAGAAAC 18077
OY 478 GCGAGGCTGTGTCTGTCGAGATGCGATGTCGACCGCATCTATGACACGAGCGCCCT 537
DB 18076 GATGCTTGATGAGCGATCCGCGATCGGCTGTGCGCATGATGATCCCGCAGATGTTGG 18017
OY 538 CGGCTCTACAGTGTGTCGCGCTTGGCTGAGACGAGCGGCAAAACCTGCTGTCGTCAGT 597
DB 18016 CGGCTTGTGATTTAGTTATTCGCGCTTGACAGGCGCAACATGCGCTTGCAATTTGAGT 17957
OY 598 ATGCATCTCATTAAGTTAGCTAGGACGACCTGTGCATCATCTTCAAGAGTGGCTACG 657
DB 17956 CTGCATCTGCTGATTCGCGATGCTCTGATGATGCGCACGCTACAGAGAGTGTATGATG 17897
OY 658 TTCTACGAAGATCCCGAGACCTCTCTCCCTGTCCTGAGCTCTCGTACCGCGATTATGTA 717
DB 17896 CTGTACCGGTGAGCCCATGTGTGCTACGCGTGTGTAACGTTCTCTTTCGTGACTACG 17837
OY 718 CTCGCGTGTGAGTCTCGCAAGAAATCTGAGCGCATCAACGATCGATGATTACTGGAAG 777
DB 17836 CAGGCGTGTGTGTAGAGCAGGCGAGTGAAGCTTATGACGCGATCAGGCGCTATTGGCA 17777
OY 778 CGGCGCATCGCGAGCTCCACCTCGCGGACGCTTCGATGAAAGCGGATCCATCTAC 837
DB 17776 CGGCGCTGCGCAGCTGTATGCGCCACACACGCTGCGGTACAGAGGATTTGGCGCA 17717
OY 838 CTGAAGAGATTCGCTTCGCGACACAGAGCAATGCTGCGCTCGAGACTCTCGGGGTGCA 897
DB 17716 CTGTGTGAGATGAGTTGTGTAGTGTGCGCGTATCGGCTGTACGCCCAACACTGGGAGTG 17657
OY 898 TTGAAGCGCGGTGTGCGGAGCGCGGCTGACCCCGACGCGGCTCATCTGTGCTCATTT 957
DB 17656 CTGAGCGGCTGCGCCCAACGCAACGTATCACCAAGACGCAATTTGTGTGACATCTTT 17597
OY 958 TCCGAGGATCGGCGGCTGAGAGCGGAGCCCGGTTTACGCTAACATAAGCTCTTC 1017
DB 17596 AGCCAAAGTGTGCGACGTTGAGCCTTACCCGACGTTTACGCTAACATGTGAGTGTTC 17537
OY 1018 AACCGGCTCCCGTTCATCCGCGGTGAACGATATCACCGGGGACTTACAGTGTGATC 1077
DB 17536 AACCGCGCGAGGGTTACCCCAACGAGAGGAGTATTGTGTGATTTAACGCTGTACG 17477
OY 1078 CTCTGTGACATGACACCATCTGCGACAAAGACTTGCAGACGCGGCTTAAGGTATCA 1137
DB 17476 TTGCTGAATGTTTGTTAAGACAGCGACGACTTATGCGCACAAAGCTCAGGATTCAG 17417
```

```
OY 1138 GAGCAGCTGTGGGAGAGCGATGATCATGCGACGTAAACGGTAT-----CGAGTTC 1188
DB 17416 GTGCAACTGTGGGAAAGATCTGAAACATGCTGTGTTTACGTGGGATCCGCGCACGAGCG 17357
OY 1189 CAGCGAAGAGCGCCCGGCTCTGGGATCCAAAGAGCGCATTTGTTCCCGTGTGCTC 1248
DB 17356 CTGATCCATTAACCGGTCTGTTTCCATGCGCGGATGCGGATGCTGATGATGTTGAT 17297
OY 1249 ACAGCGCGCTTAAACAGCAAGTGTGTGTCACCTGCTGTGACAGAGCTGGAACCTCG 1308
DB 17296 ATCGACGGGAGAGAGACTGCGCAAGACCTTGGGACACAAACCTTTTACTGTGTGCG 17237
OY 1309 GTGTACACAGACAGACAGACTCTCAGCTGCTGTGATCATCACTTACAGTACAGACAT 1368
DB 17236 GAGCGCAATATTATCCCAAAACCGGAGGTGTGCTGATCATCAAGTATGAGTGTGCT 17177
OY 1369 GGGGACTCTGTCCTCGCGTGGACATGTCGACGAGTGTTCGCGCCGACCTTGTGAC 1428
DB 17176 GGGGAGTTGCAATTCACTGGGACGCGGTGAGCAACTGTTGTATACCAACGCTGTGAT 17117
OY 1429 GACATGCTCGAAGGATACGTGTTTCTCGGCGGCTCACTGAGGAACCATGGGTAA 1488
DB 17116 CAGATGTTGTGCTTATTTGATGCTGCTGAGCGCTGTGTCATGCC-----G 17066
OY 1489 CAGGTGCGCTTTCGCTTCCGCTGCGCAGCTAGAAGCGCGGAGCGCAACGCGAC 1548
DB 17065 CAAGTGTGTGGGGGTAAATAGTTCTGTGGCGCTGCCACCGTTAAGTACACCGGTACG 17006
OY 1549 AACGCGCTGTAGACAGACATACGTCGACGCGCTGTTTCGCGCGGGGTGAGCAGT 1608
DB 17005 CAGGCTCTGACCTACGCGCCTTGTGTCACATGAGATTAACGTCAGCAGACATGACG 16946
OY 1609 CCCATGACGCTGCGCTGTGTGTCGCGCGGCAAGACGCTCAAGTACGAAGCTTTGGGCG 1668
DB 16945 CCACAGAACTGTGCGTATAGTCTATCCGTGAATGACTTATCGCCACTGTGACG 16886
OY 1669 GCTTTCGCGCAGCTTGTGCGCGCGCTGCGCAGAGAGGGGCGACGCGCCAAACATTTGTC 1728
DB 16885 GCGGCGATCATGTGTGCGCGCGCTGTTTACGCGTGGCGTGTGACATGGGACCGCGTG 16826
OY 1729 CGGCTGTGATGAGAAAGCTGAGAGAGGTTGTCCGCTTCTGCGGTGCTGAGTCA 1788
DB 16825 GCGGTGTGATGAGAAAGCTGAGAGAGTTCGCGCTGACAGTTCGCGCTTATACGCTG 16766
OY 1789 GCGCGGCTACGTCGCGATGATGATCCGACCTACCGCGGAGCGATCACTACTCTC 1848
DB 16765 GGTGCGCTATCTGCGACGTGATCCGCTGCTACCGCACAGCGTCCAGCTTTTGTG 16706
OY 1849 GATCATGATGATGAAGCTGTGCTGACGAGCATGCTGTGATGCAAACTGTCATG 1908
DB 16705 ACGGTGGCGAGGTGCGGTCAAGTATACGACGCGGCTTCAAGCAATTGAGCGCTG 16646
OY 1909 CCGCGGGGATCCAGCGGCTGCTGTGAGCGAGCGCGGCTGGAAGGCGAGCGACAG 1968
DB 16645 CTGCGCG-----TGTGATCATGACAGCGGAATGTGGAACACCGCTGTGCG 16598
OY 1969 CCTCGAGTATGCCATTTCAACACCTTTCGATCTCGGTATGTCACTTACACCTCGGGA 2028
DB 16597 CCGTTCGTGAAGTGTGCGGAGATGTCAGGATCTGCGCTTATCATTTTACATTCGCG 16538
OY 2029 TCCAGAGGTTGCCCAAGGGGATGATGATCATCATCGGGGTCGTCACACCATCTCTG 2088
DB 16537 TCCACCGTATCCCGAAAGAGTATATATCAACACCTGCGGCTATTAACACGCTGGAA 16478
OY 2149 AGCTTCAATCTCTGCTATGATGTGTGCGGATCTCGGCGGGGCGGTATGATCTG 2208
DB 16417 AGCTTTCAGCTGTGTTTACGATGCTTTCGCGCTTTTATGTGTGGTGAAGCGCTGTA 16358
```


QY 2209 GTGCCGAGACCGCTCAAGCTGCGCCGATTCGGCGCAGATTGGACAGATTGACACGAG 2268
DB 16357 CTGCCGGAAGCAGAGACGGAAGAAAGATCCGCTGATTGGCAGACAGTTATGACACGAT 16298
QY 2269 AAGGTGACGGTGTGAACTCGGTGCGGCGCTGATCGGATGCTGTCGAGCATTTTGA 2328
DB 16297 CATGTAAAGGTCTGGAATGCAAGTCCCGACATGATGATGCTGCG---AATACAC 16241
QY 2329 GGTCCGCCGATTCGCTCGCTAGAGTCTGCGGCTTTTCGCTGTGAGCGCGACCTGATC 2388
DB 16240 ACCGCGCATCGGATGAGTTATCCGACGTTGCGCTGAGCACTGTGAGCGCGCATGATC 16181
QY 2389 CCGGTGGGCTGCTGCGCA---GTCACGCGCATAGAGCGCGCTGTCGATGATCAGC 2445
DB 16180 CCGCTAACGTTACCGAGCAGATGCGCAGCGGCTCAATGAAACGATGACATCATCAT 16121
QY 2446 CTGGCGGGGACCAACGAGCGTGCATCTGGTTCATCGGGTACCCCTGAGAACTGCAC 2505
DB 16120 CTGGGTGAGCAGACCGAGTCCGCACTGGTCCGCTTACTACCCGATAGGTGAGTGA 16061
QY 2506 CTATCTGGGCGAGCATCCCTTAAGCCGCTGCTGCGCAACGACGTTCACTGCTC 2565
DB 16060 TCGACGTGACACGATTCCTTACGCTGCGGCGCTGCGCAACGACGATGATGCTTA 16001
QY 2566 GATGAGCCCTGCAACCGCGCCGCTGCTGCTGCTGCGCACTTACATTTGGCGGGCTC 2625
DB 16000 AATGCCAACCTGAGGAATGTCGCTGCGGCTGGAAGAGATTTGCAATGCGGGAATG 15941
QY 2626 GGGCTGGCACTGGCTACTGTCGCGATGAAAGAGAGACGCAAGACTTCTCTGTCAC 2685
DB 15940 GGGCTGGCAAGGCTTACTGTCGCGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 15881
QY 2686 CCGGAGACCGGGGAGCGCTCTTAAAGACCGCGATCTGGGCGCTTACTGCGCGATGA 2745
DB 15880 GAAGCAGAGTGTAGGGAATTTTACCGCACTGCGGATGCGCGGCGCTTACTGCTGA 15821
QY 2746 AACATGAGATTCATGGGCGGTGAGCAACCAATGAGCTTCCGCGATACCGCGTTGAG 2805
DB 15820 CAAGTGGCTTTTGGGCGCAACGATCCCAAGTGAAGTGAATGATGATGATGAA 15761
QY 2806 CTGGGGAATCGAGAAACGCTCAAGTGCATCCGACGTCAGCGACGCGGTGATTTG 2865
DB 15760 CTGGGGAATCGAGGCTGTCATTTGGCGCATCCGAGTGTGAGGAGAGTGTGATG 15701
QY 2866 CCGGTGGGAACGACCGCGGAAACAAAGCTCTTCTAGCTATGTGCTCCGAGGCGACA 2925
DB 15700 GCAGTGGGTAA-----TTCTCAACATCTGCTCG 15674
QY 2926 CCGAGACGCGCTGCGAGCAGACGCGAGCTCAAGCCGAGCGATTCGACGAGAGCA 2985
DB 15673 CTGGTGGCTTTTCCAACTGACAGATCCGACCGAGCGCATTTGCAAGCTTAAAGAA 15614
QY 2986 CACGCGCGAAGCGAGCGCTTGAAGCAGCGGAGAGGAGTTCAGTCAAGCTCGCTCA 3045
DB 15613 -----GCGAGCGCGGCGGACCTGGGCGAGGATTAATGTGAATCCGAGCAGCGCTA 15560
QY 3046 CACGAGCTCGGAGGAGCTGAGCGGAAACCGGTGCTGATTTGACCGGAGAGATCCG 3105
DB 15559 GCGTTCAACTCAAGGAGCACAATTCGCGCGCTGATGCTGCGGCAATTCGATGAC 15500
QY 3106 CCGGAGCGCGGCGGTGAGCGTCAAGCGCGCTGAGCGCTGCGAAAGTTCTTGAAGCC 3165
DB 15499 GCAACCGCGATTCACACGTTTCAATCAACGCGGAGCTATGCTCATTTTCAAGCGGCA 15440
QY 3166 CCGATTCCGTTTGTGAGTGTGCTGATTCCTGAGCTGCTTGAAGCGCTGAGCGCGAC 3225
DB 15439 AAAACACGCTGACAGAGTGGGCAATTCCTGCGGCTTGGGCGAGAGCGCTTACCC 15380
QY 3226 GCGCGGACCTTCCCAATTCCTTATCATTCGCGGCGGACGACGTAACCGGTGCAACC 3285
DB 15379 GGTCTACTTTTTCGCAAGTATGCTTATGCTTCCGCGCGGCGGCTTATCCGCTGCAACC 15320
QY 3286 TAGCGGATGTCAATTCGCGCGCATCGAGGCGGTGAGCAGAGGCTTCTATTTATCCAC 3345

DB 15319 TACGTATCTGATTCACAGACAGATCGAAGAGGAGATTCGGATTTTACTTTCAC 15260
QY 3346 CCGTTGAGCAACGTTTCTGAAGCT---CTCCATCAACGAGATCGAGCGCGAGCGAC 3402
DB 15259 CCGGACAGAGCTGCTTATGCGGATACACAGAAATCGAGCTGAACAGTGTGTTTAT 15200
QY 3403 GTTCCGCAAACTTGAAGTGTTCATGAAAGGGGTTCAACCTCTCTGTTGCGGACG 3462
DB 15199 GCGGACCTTAACAGTCAATGCTGATGCGGCGGATTCACGCTTTATGCTGCTGAT 15140
QY 3463 ATGACGCGATGATGCTGCTGATG---ATGCTGTCGCGAGAAATTTTGCCTGAGAG 3519
DB 15139 ATGCGGATGATTCGCACTTCTATGCGGAGAGAGAGAGCTTCTGCTGATGAA 15080
QY 3520 GCGGATATATGAGCGAGCTCTGATGAGACAGCGCTTCTCTGCAACATCGGCTGCT 3579
DB 15079 GCAGGTACTCTGTCATTTACTGGAAGAGATGCGCGCTACGATGCGAGTGGGCTGTC 15020
QY 3580 CCGGTGGGCAATTCATTTTGAACAGTTCGCGCGGTTCTGACCTGCGACATTCGAC 3639
DB 15019 CAATTTGGATGAGCAGACTTTCGCTGCGGATTCGATTCATTTGCTGCGACATCAT 14960
QY 3640 GTTTACGTCACGCGATGCTGGCGGCGGATGACCCGCGCACTTCAGGTCTGTACG 3699
DB 14959 CCGTATGTCATTTGACCGCTGGGCGCGATGAGGCAAGAGCGGCAAGTCTTCACGA 14900
QY 3700 CTGGTCAAGATTCCTCAACGAGCGCGCACGACGCGCGCGCTTCCGCGCGAG 3759
DB 14899 TTGCTGCGGATTTCTCCACCTATGAGAAACGAGAGAAACCGCTGCGCGCTGACATG 14840
QY 3760 CAGCACTTGCAGATATGCTTTCGCACTTTCTTGAAGACCAATCCGAGTACATG 3819
DB 14839 CAGACTTAAGAGATGCACTGCTGCGCGCTGCTGACAACTGCTGATTAATG 14780
QY 3820 CTTACAGTCTTGTGAGAGCTGATGCTGCGCTGACGTCGCAAGCGAGTTCATGCT 3879
DB 14779 CCGAGTATGATATGTTAGCAGACCGATTTTCCGTTAACCGCTAACGAGCTAATG 14720
QY 3880 AAGGCTCGCGGAGGAGATCTGCTGCTGCGCGCAT-----TCGGGCGACAG 3933
DB 14719 CAAAAATTAACACTCAGGCGCAAAATTCGCCACAGCTGACGCGCTGGTCCATC 14660
QY 3934 GCGGACGAGGAGCTTGGAGAGATTCCTGCGGCTGCTGCTGAGAGGTGCTCGGCG 3993
DB 14659 CAGGTGACAGTGCCTTACACAGCGCGCTGCTGAGGCTTGGCAGAGATCTCGCG 14600
QY 3994 GAGTGTGCGGCTCAGCAGAGCTTCTGATCTTGTGCGACATTCATTCATCTT 4053
DB 14599 AGCAGCTGTGCGGCGGAGAGCATTTTCTGCTGCGGCGGCGAGTTCTATAGATG 14540
QY 4054 CCGATGAGAGCTGTTGCAAGAGGCTGATAGGAGATGCGATCAGCGATTTTTC 4113
DB 14539 CCGATTAAGCAGGACCTTGAAGCGATTTATCGGCGAGAGATTCATTTGCTGATCT 14480
QY 4114 CAGTACCGAGACTGCGCTGCTGCGG 4140
DB 14479 CGTCTGCCAACATCGCGGATGTGG 14453

RESULT 6
ABST8890/C
ID ABST8890 standard, DNA; 50538 BP.
XX
AC ABST8890;
XX
DT 17-DEC-2002 (first entry)
XX
DE E. coli CT073 genomic sequence #57.
XX Pathogenic; Escherichia coli CT073 infection; livestock; pyelonephritis;
XX urinary tract infection; open reading frame; ORF; uropathogenic;
XX antibacterial; aropathic; nephrotropic; gene; de.

XX Escherichia coli.
OS
XX
PN W020259320-A2.
XX
PD 01-AUG-2002.
XX
PF 19-OCT-2001; 2001WO-US46833.
XX
PR 19-OCT-2000; 2000US-242412P.
XX
PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX
PI Blattner FR, Welch RA, Burland VD;
XX
XX WPI, 2002-691532/74.
XX
XX New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
PT useful for preventing or treating E. coli CFT073 infection in humans or
PT livestock -
XX
XX Claim 1; Page 186-214; 765pp; English.
XX
XX The present invention relates to polynucleotide sequences from the
CC genome of the pathogenic Escherichia coli strain CFT073. Almost all
CC the sequences present in E. coli CFT073 are absent in the previously
CC sequenced laboratory strain K-12. The polynucleotide sequences of
CC the invention are useful for preventing, diagnosing or treating
CC E. coli CFT073 infection in humans or livestock. The polynucleotide
CC sequences are useful for preventing urinary tract infections and
CC pyelonephritis. Likewise, the polypeptides encoded by the different
CC open reading frames (ORF1-5) are useful for generating a vaccine
CC against uropathogenic E. coli strains. ABS7834-ABS7905 represent
CC genomic sequences from E. coli strain CFT073.
XX
XX Sequence 50538 BP; 12312 A; 13865 C; 12701 G; 11654 T; 6 other;
SQ
Query Match 12.8%; Score 543.4; DB 24; Length 50538;
Best Local Similarity 48.5%; Pred. No. 8.5e-106;
Matches 1933; Conservative 0; Mismatches 1961; Indels 93; Gaps 12;
QY 181 ATCTGCCCCCGCCAGCGAGCGGACGCTCCGTTCTCTCAGACATCCAGAAATCC 240
DB 20201 ATAGGACAGACCCCTGAGCATGTCATCAACCGTTCCCATTAAGCATGTCAGCGCC 20142
QY 241 TACTGGCTGGCGGCGACAGAGCTTTACGTTCCCGAGCGGATCAGCCCTATGCGCA 300
DB 20141 TACTGGCTTGAACG--ACAGACGGGGTCTACCTGATCGGACCCACATTAACATGAA 20085
QY 301 TACGACTGTACGAGATCTGACAGCTGCGAGGCTGACCGGCTTTGCGAAAGTCTGCG 360
DB 20084 TTTGACGTAGAAACATTTATATTTACGGCTTTACCCATGCGGTAAATCCGTATCCG 20025
QY 361 CGGACGACATGCTTGGGCGCCACACGCTGCCGACATGATGACAGTATGAGCTTAA 420
DB 20024 CGCCATGAAGTCTAGTGGCGGGTACTCCCGACGTAACAGAGATTTGCGGAA 19965
QY 421 GTCAACGCG--CGAATGAGATCATCATCTGCGCGGGCTGACCGGAGCACAGGAA 477
DB 19964 GTGCGGGGTATCACTTATGAGAGCGGATGATGCTTTGTCCCTTAAACGACGAAAC 19905
QY 478 GCGAGGCTGATGCTTGGAGAGATGAGATGTCGACCGCATATATGACACCGAGCGCCT 537
DB 19904 GATGCTTGAATGAGCATCCGCATCGGCTGTGCAATATGCAATCCCGAGATGTTGG 19845
QY 538 CCGCTTATCAACGCTGCTGCGCGCTGAGACGAGCGGAAACCGCTCTGCTCACT 597
DB 19844 CCGCTTATGATTTAGTTATTCGCTTGAACGCGGCAACATGCGCTTGAATTCAGT 19785
QY 598 ATGATTCATTAAGTTGACCTAGGACCGCTGTCATCATTTCAAGATCGGCTCAGC 657
DB 19784 CTCGATCTGCTGATGCGGATGCTCTGATGATGCGACGATACAGAGAGTTGATGATG 19725

QY 658 TTTCAGGAAGATCCCGAGACCTTCTCCCTGCTGAGCTCTGCTACCGGATATGTA 717
DB 19724 CTGTACCGTGAAGCCCGCATGTGTCACTGCGTTGATACGTTCTCTTTTGTGTATCAG 19665
QY 718 CTGCGCTGAGCTCTCGCAAGATCTTGAAGCGCATCAACGATGATGATTTACTGAA 777
DB 19664 CAGGCGCTGTTGTAGAGACGAGCGGATGAACTATGACCGCATCAGGCTTATGGCA 19605
QY 778 CGGCGCATGCGCGGATCCCACTTCGCGGACGCTTCCGATGAAGCGGATCCATCTAC 837
DB 19604 CGGCGCTGCGGACGCTTATGAGCCCAACGCTTCCGATCAGGCGGATTTGGCGCA 19545
QY 838 CTGAAGGAGATCCGCTTCCGACACAGGAGATGCTCCGCTGAGCTCTGAGGTGA 897
DB 19544 CTGTCTGGATCAGGTTGTATGCTGCGGCTATGCGGTGTGACGCCACAACTGGGAG 19485
QY 898 TTGAAGCGGCTGTGCGGAGCGGCTGACCCCGAGCGGCTATCTGCTCATTT 957
DB 19484 CTGAGCGGCTGCGGCGCAACGACATATCAACAGAGGATTTGTCACGCTCTT 19425
QY 958 TCCGAGTGTATCGGCGCTGAGCGCGGACCGCCCGTTTACGCTCAATACGCTCTTC 1017
DB 19424 AGCCAGTGTGCGGACGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTTC 19365
QY 1018 AACCGGCTCCCGCTTCATCCGCGGTGAACGATATCACCGGAGCTTCACTGATGTC 1077
DB 19364 AACCGGCTCCCGGCTTACCCACGACAGGAGATTTATGATTTTACCGCTGTACG 19305
QY 1078 CTCTGACATTCACACCTTCGCGACAGAGCTTTCAGACCGGCTTAAAGGTATTCA 1137
DB 19304 TTGCTGAATGTTTGTATGACAGCGGACCTTATGCGCCACAGCTCAGGATTTG 19245
QY 1138 GAGCAGTGTGGAAGGAGATGATCATGCGACGTAACGCTAT-----CGAGGTC 1188
DB 19244 GTGCACTGTGGAAGATCTGAAATGCTGTTTCACTGAGATCCGCGCACGAGGCG 19185
QY 1189 CAGCGAGAGCGCGCGGCTCTGCGGATTCACAGAGCGGATTTTCCCGTGTGCTC 1248
DB 19184 CTGATTCATTAACGCGGCTTTCATATGCGCGGATTCGATTAATGATGTTGAT 19125
QY 1249 ACGAGCGCTTAAACAGCAAGTGTGTACCTGTTGCAAGGCTGGAACCTCG 1308
DB 19124 ATCGACGGGGAGAGACTGCGCAAGACCTCGGAGACACACCGTTTATCTGTGTCG 19065
QY 1309 GTGTACACAGCAGGAGATCTCAGCTGCTGCTGATCATCACTTACGATGAGAGCAT 1368
DB 19064 GAGCCATATTTAACCAACACCGAGGTGCGTCTGCAACAGATATGAGTGTGCT 19005
QY 1369 GGGGACCTGCTCTGCGCTGGAACATCTGACGAGAGTTCGCGCCGACCTTGTGAC 1428
DB 19004 GGGGAGTTGATTTCACTGGAACGCGGTGCAAGCACTGTTGATACACGCTGTGAT 18945
QY 1429 GACATGCTGAAGCTATGCTTTTCTCCGCGGCTCACTGAGAAACATGGGGTAA 1488
DB 18944 CAGATGTTGCTGATTTATGCTATGCTGAGCGCTGAGGCGCTGTGGCCATCC-----G 18894
QY 1489 CAGTGCCTGTTGCTTCCGCTGCGGCTGAGTAAAGCGGCGGAGGCAACGCGAC 1548
DB 18893 CAAGTTGTGAGGGGTAATATGTTCTGTGCGCTGCGACCGTTATGTCACGCTCAC 18834
QY 1549 AACGCGCTGTAGAGACATACCTGACGAGGCTGTTCGCGCGCGGCTGAGCAGCTG 1608
DB 18833 CAGGCTCTGACCTAGCGGCTTGTGTGACCATGATTAACGCTGACGACATGACG 18774
QY 1609 CCATGACGCTGCGGCTGTGCGGCGGCAAGCTTCACTGATGAGAAAGCTTTGCGGC 1668
DB 18773 CCACAGGAACTGCGGCTGATATGCTATCTGTAATGATGATCTATCGCAATGTCG 18714
QY 1669 CGTTGCGGCGGACCTTTGGCGCGGCTGCGAGGAGCGGCGGACGCGCAATGCTG 1728
DB 18713 GCGGCGATCATGTGCGCGCGGCTGTTACGCTGCGGCGGTCGACATGCGACCGCTG 18654
QY 1729 GCGGTGTGATGAGAAAGCTGAGGAGGCTGTGCGGCTTCTGCGGCTCTGAGTCA 1788

18653 GCGGTGTATGAAAAAGCTGGCAGAGTTGCCGCTACCGCATTTTACGACTG 18594
1789 GCGCGGCGCTACGTCGATGCTGATGCGGACTACCGGCGAGGATATCCACTCTC 1848
18593 GGTGGCGTTATCTGCACTGATGCTGCGGTACCGGCAAGGCTGCGAGCTTTGCG 18534
1849 GATCATGTGAGTAAAGCTCGTCTGACGCGGCAATGCTGATGCGAACTGTATG 1908
18533 ACGGTGGCGAGGTGCGGTACAGTACGCGCGGTCTCAGCAATTTGAGCCGTG 18474
1909 CCGCGCGGATCCAGCGGCTCGGAGCGGCGGCGGATGAGCGGCGAGCGAG 1968
18473 CTGCGCG-----TGTGATCATGACGAGCAATGCTGACACGCTGCTGG 18426
1969 CCTCCGATGATGCCATTCAGACACTTCGATCTGCGTATGCTACACTTCGGA 2028
18425 CCGTTGCCGAGTGGCTGGGATGTCAGGATCTGGCTATATCTTTCACTTCGGC 18366
2029 TCCACAGGTTGCCAAGGCGGTGATGATCATCGGCGTCCGTCACACCATCTG 2088
18365 TCCACCGGTACCCGAAAGAGTATGATGACACCGTGGCGCATGAAACGCTGGA 18306
2089 GACATCAAGAGGCTTCGAAATAGGCGCGGAGCAGGATGCTGGCGCTCTGCG 2148
18305 GACATCAAGAGGCTTCGAAATAGGCGCGGAGCAGGATGCTGGCGCTCTGCG 2148
2149 AGCTTCGATCTCTCGTCTATGATGCTGCGGATCTGCGCGGCGGCTGATGATG 2208
18245 AGCTTCGATCTCTCGTCTATGATGCTGCGGATCTGCGCGGCGGCTGATGATG 18246
2209 GTGCGCGAGCGCTCCAGCTGCGGATCCGCGCATTTGGCGAGATTGATCGAAGAG 2268
18185 CTGCGCGAGCGAGCGGAGAAAGATCCGCTCATTTGGCGAGCATTTGACACG 18126
2269 AAGGTGAGGTGAGAACTCGGCGCGGCGGCTGATGCGGATGCTGCGAGATTTGAG 2328
18125 CATGTAGGCTCTGAGATGAGTCCGCGCATGATGAGTGTGCG---AATACAC 18069
2329 GGTGCGCGCGATTCGCTCGCTAGTCTCTGCGCTTTGCTGCTGAGCGGCGAGT 2388
18068 AAGGCGGATCGAGTGAATATCCGAGCTTGGCTGCGCATGTTGAGCGGCGAGT 18009
2389 CCGGTGGCGCTGCTGCGCA---GCTCCAGGCGCATCAGGCGGCGCTGCGATGAC 2445
18008 CCGCTAACGTTACCGGAGCAGATGCGGCGGCTCAATGAACATGACATCATCA 17949
2446 CTGCGCGGCGGCGAGCGAGGCTGATGCTGCTGCGGATCCCGGAGAGAACTGAC 2505
17948 CTGCGGTGAGCGAGCGAGTCCGATCTGCTGCTACTACCGGATGAGTGAAG 17889
2506 CTATCTGCGGCGAGCATCCCTTACGCGCGCTCGCTCGGACCGAGCTTCCAGCT 2565
17888 TCGAGTGTGACCGATTTCCCTTACGCGGCTCGGCGGCGGAGATGAGTGA 17829
2566 GATGAGGCGCTGACACCGCGCGCTGCTGCGGTTCCGCGGCACTTACATTTGCG 2625
17828 AATGCCCACTGAGAGATGTCTGCGGTGAGAGAGATTTGCAATTTGCGCGGAG 17769
2626 GGGCTGGCACTGGGCTACTGCGCGGATGAAAGAGAGCGCAAGCTTCCCTGAGC 2685
17768 GGGCTGGCAAGGCTTACCTGAGAGCGAGAGAGAGAGCGGCGGAGCTTTGTCT 17709
2686 CCGGAGACCGGAGCGCTCTTACAGACCGCGGATCTGGGCGCTTACCTGCGAT 2745
17708 GAAAGAGGTGTAGGAGATTATCCGACCTGGGAGTCCGCGGCTTATGCTGAG 17649
2746 AACATGAGTTCACTGGGCGGTGAGACCAACCAATCAAGCTTGGCGATCCGCT 2805
17648 CAAGTGCCTTTTGGGCGGCAAGATACCAAGTGAAGTGTATACCTATCGAA 17589
2806 CTGCGGGAATGAGAGAAAGCTCAAGTGCATCCGAGCTAGCGCGGCGGATTTG 2865

17588 CTGCGGGAATTCAGCGCTGCAATTGCGGACATCCCGATGTGAGACAGTCA 17529
2866 CCGTGGGAGACGAGCGGCGGAACAGCTCTTACCTATGTGTCCGAGGGCACA 2925
17528 GAGTGGGTAA-----TTCTCAATGTGTGCG 17502
2926 CCGAGACCGCTGCGGAGAGGAGCGGAGCTCAAGACCGAGCGATGACGCGAGAGA 2985
17501 CTGTGCTTTTGGCAAACTGACAGATCGCACAGGCGGAGGATTTGCAACTTA 17442
2986 CAGCGCGCGAAGCGGAGCTTGAAGCAGCGGAGAGGAGTGCAGTTACGCTCGA 3045
17441 -----GCGAGGCGGCGGACATGGCGAGGATATTTGTGAATCCGCAAGG 17388
3046 CAGGACTTCGAGGAGACTGGAAGGAAAGCCGTGCTGATCTGACGCGGAGATCG 3105
17387 GCGTTCAACTCAGAGACCAATATTCGCGCGCTGAGTGTGCTGGGATTCAGT 17328
3106 CCGGAGGCGGCGCTGAGAGTACGCGGCTGCGGCTGAGCGGAGCTTGAAGCC 3165
17327 GCACCGGCGGATGACACAGCTTACAAACGCGCGAGTATGCTATTCAGCGG 17268
3166 CCGATTCGCTTGTGAGTGTGCTGAGTCTGAGCTGTTGAGCGGCTGAGCGG 3225
17267 AAACACGCTGGCAGAGTGGGCAATTCGTGTGGGCTTGGCGAGATGCTTACC 17208
3226 GCGGAGCCCTTCCAAATTCGTTATCATGCGGCGGAGCAGTACCGGTGCAACC 3285
17207 GGTTCACCTTTTCCAGATATGCTATGCTGCGCGGCGGCTTATCCGCTGAA 17148
3286 TACGATATGCAATTCGCGCGGCGGCTGAGCGGAGGCTTATATTAACAC 3345
17147 TACGATATGCAATTCGCGCGGCGGCTGAGCGGAGGCTTATATTAACAC 17088
3346 CCGTTCGAGACCGTTCGCTGAGCT---CTCCGATCAAGGATGAGCGGCGG 3402
17087 CCGGACAGAGTGTCTTATGCGGATGACAGAGTGCAGTAAAGTGTGTTAT 17028
3403 GTTCGCAAACTTCGAGCTGTTGATGAGGCGGCTTCACTCTGTTGCGGAG 3462
17027 GCGGACCTTATCAGTATATTCGCGGCGGAGTTCACCTGTTATGCTGAT 16968
3463 ATGAGCGCATGAGTGTGCTGATG---ATGCTGCGCGGAAATTTGCTGAG 3519
16967 ATGAGGATGATTCGCTATTCATGAGGAGAGGAGCTTGCATCTTCGATGA 16908
3520 GCGGATATATGCGCAGCTCTGATGAGAGGCGGCTTCTGCAACATCGGCTGT 3579
16907 GCAGGACTCTGCAATTTATGGAAGAGATGCGCGCTACGATTTGGGCTGT 16848
3580 CCGGTGGGCAATTAATTTGAACAGTTCGCGGCTTCTGACCTGCGAATTCGAG 3639
16847 CAATTTGGGATGCAAGATTTTCGCTGCGGATGCAATTTCAATTTGCGCACT 16788
3640 GTTACGTCAGGATGCTGCGGCGGAGGAGTACACCGCGGAGTTCAGGCTGTAG 3699
16787 CCGTATGCTATGCACTGAGCGGTGCGGCGGAGTACAGGAGGCGGAGTCTGAG 16728
3700 CTGCGTCAAGATTTCTCAACGAGCGGCGGAGGCGGCGGCTTCCGCGCGGAG 3759
16727 TTGCTGCGGATTTCTCCACTATGAGAAACGAGAGAAACGCTGCGCTGAG 16668
3760 CAGCACTTTCGAGATGCTTGGCGACTTTTGAAGACCAATCAACCGAGTACAT 3819
16667 CAGAGCTACAAAGATGCTGCTGCGGCGGCTGCGTCAAGATCTGCTATGAG 16608
3820 CTAAGCTTTCGAGAGTCTGATGCTGCGGCTGAGCTTCCAGCGGAGGCTGAT 3879
16607 CCGAGTATCTATGTTAGCGACCGATTTTCCGTTACCGCTTACGCGAGCTAG 16548
3880 AAGGCTTCGCGGAGGAGATACCTGTCGCGGCGAT-----TGGGCGACAG 3933
16547 CAAAATTAACCTGAGGCGGAGCAAAATTTGCCACACAGCTGAGCGGCTGAGT 16488

QY 3934 GCGCCACGGAGCGCTTGGAGAGATCTCTCGTGGCTGTACGGAGGTGCTCGGGCTG 3993
DB 16487 CAGGTGAGCAGTGGCTTACACAGCGGCTGGCTCTGCGACAGAGTACGGCGCTG 16428
QY 3994 GAGGTGCTGGGCTTCACAGAGCTTCTGTCATCTTGGTGGACATTCATTCATCTGTT 4053
DB 16427 AGCCACGTGTGCGCCGAGACGATTTCTCTGCTGGGGGCGCATTTATAGAAATTGGTG 16368
QY 4054 CGCATGAGAGGCTTTCGACAGAGGCTGGATAGGAGATGCGCATCAGCGAGTTGTC 4113
DB 16367 CGTATTCAGCAGGCACTTGAGGCGATTCGCGCAGAGATTCCTCATTTGTGATCTGTC 16308
QY 4114 CAGTACCGGAACCTCGGCTCGCTGGCG 4140
DB 16307 CGTGTCCACACATCGCGGATGTGGCG 16281

RESULT 7
ABST7887
ID ABST7887 standard; DNA; 32160 BP.
XX AC ABST7887;
XX DT 17-DEC-2002 (first entry)
XX E. coli CFT073 genomic sequence #54.
XX Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
KW urinary tract infection; open reading frame; ORF; uropathogenic;
KW antibacterial; atropathic; nephrotropic; gene; ds.
XX OS Escherichia coli.
XX EN WO200259320-A2.
XX 01-AUG-2002.
XX 19-OCT-2001; 2001WO-US46833.
XX PF 19-OCT-2000; 2000US-242412P.
XX PR 19-OCT-2000; 2000US-242412P.
XX PA (WISC) WISCONSIN ALUMNI RES FOUND.
XX PI Blattner FR, Welch RA, Burland VD;
XX WPI: 2002-691532/74.
XX PT New DNA sequences of the pathogenic Escherichia coli CFT073 strain,
PT useful for preventing or treating E. coli CFT073 infection in humans or
PT livestock
XX Claim 1; Page 166-184; 765pp; English.
XX The present invention relates to polynucleotide sequences from the
CC genome of the pathogenic Escherichia coli strain CFT073. Almost all
CC the sequences present in E. coli CFT073 are absent in the previously
CC sequenced laboratory strain K-12. The polynucleotide sequences of
CC the invention are useful for preventing, diagnosing or treating
CC E. coli CFT073 infection in humans or livestock. The polynucleotide
CC sequences are useful for preventing urinary tract infections and
CC pyelonephritis. Likewise, the polypeptides encoded by the different
CC open reading frames (ORF1-5) are useful for generating a vaccine
CC against uropathogenic E. coli strains. ABST78834-ABST79085 represent
CC genomic sequences from E. coli strain CFT073.
XX
SQ Sequence 32160 BP; 7001 A; 9686 C; 8766 G; 6688 T; 19 other;

Query Match 9.2%; Score 390.4; DB 24; Length 32160;
Best Local Similarity 51.2%; Pred. No. 3e-73;
Matches 1086; Conservative 0; Mismatches 997; Indels 38; Gaps 6;

QY 761 CGATGATTACTGGAAGCGCGCATTCGCGAGCTCCACCTCGCGCGATTCGATGA 820

DB 10021 GCGCGCTTACTGTGCTGGCGAAGCATGACGCTTCCCGCGCGCTTCTGCGCTGG 10080
QY 821 AGCGCATTCATCTAACCTGTAAGAGATTCGCTTCCGCGACACGAGCAATGGCGCGT 880
DB 10081 CTTGGAAACCGCCACGCTATATGTAAAGTCTTAATACCCAGCCCGCATGATTTGCGG 10140
QY 881 CGSACTCTGGGGTTCATTTGAAGCGGCTGTGGGAGACCGGCGCTGACCCCGAGCGGCG 940
DB 10141 CAACACGTGGCAGCGCTTTATAGCAACCGGGCGGGAGATATAGCGGTGACCGCAATGG 10200
QY 941 TCATCTGGCTGCAATTTTCCAGAGTATGGGCGCTGGAGCGGACCCCGGTTTACGC 1000
DB 10201 CGCTGGCACTGTTTTTCTCCGCTGGCTGGCTGGGGGCGGCTGACGCGCTGTGCTGC 10260
QY 1001 TCACATTAAGCTCTTCAACCGGCTCCCGCCATCCGCGGGTGAAGATATACCGGGG 1060
DB 10261 TTACATCATCTTATTTTACCGCCAGCGCTGACCGCGGATGGCGCATGCTTGGCG 10320
QY 1061 ACTTCACGTCGATGGTCTCTCTGACATCGACACCTCGGACMAAGACTTTCGAACAGC 1120
DB 10321 ACTTCACCAATATCTTCTGCTGACACCGCTGAGATGGGATACCGTCAGCAACTGG 10380
QY 1121 GCGCTAAGCGTATTCAGAGCAGCTGTGGAGAGGAGATCATCTGCACTTAACGGTA 1180
DB 10381 CGCGTAA---AAACAGCTCACGTTTACGAGAGACTGGAGACATGCGCATGTGCTGGCG 10437
QY 1181 TCGAGTTCACGAGAGGCGCGCGCGCTGCTGGGATTCMAACGAGCGCATTTGTTCCCG 1240
DB 10438 TCGAATTAATCTCGTGAATCAAGCGCAGCGCTTACCCCGCGCG---CCGCG 10491
QY 1241 TGGTGTACGAGCGCGCTTAACAGCAAGTGTGTGTGTCACCTGTTGACGAGAGCTCG 1300
DB 10492 TGGTATTTACAGCAATTTGGGGGCTTCCCTTACAGACGCGCGCAATTCGCGTTGG 10551
QY 1301 GAATCCGGTGTACACGACGACGAGACTCTTCACTGCTCTGTGATCATGACTCTACG 1360
DB 10552 GCGAGCCGGAATGGGGCATTCGCAACGCGCAGAGTGTGATAGATCATCTGCGGTTGG 10611
QY 1361 AGCAGATGGGAGACTCTCTCGCTGGGAGCATTCGTGACGAGAGTGTCCCGCCGAC 1420
DB 10612 AGCATCAGGGAGAGTCTGCTACAAATGGACACMAAGACGCGCTGTTCCTCGGCGT 10671
QY 1421 TTCTGACGACATGCTCGAAGCGTACGCTGTTTTTCTCCGCGGCTCATAGGAACAT 1480
DB 10672 TAGTCGAAACATTTGTTGACGCTTACCTCCAGTTGATTAACCACTCTGCGATGACGAA 10731
QY 1481 GGGGTGAACAGGTGC-----GCTGTTCGCTTCCGCTGCCAGCTAAGAGCGGGCG 1532
DB 10732 GCGCTGGCAAAAGCGCTTGGCAGATATATATGCGCGCAGCGAGCGCGATPAGCGAAC 10791
QY 1533 GAGCGCAAAACCGACCAACGCGCTGTGAGGAGCATACGCTGCACGCGCTGTTGCGGCG 1592
DB 10792 GGGATCAACKCCACCGGGGTGCCCCCATTCGCCAAGGCTTGTGATGAAAGCATTTTCCG 10851
QY 1593 GCGGGTGAAGAGTGGCCCATGACGCTGCGCGTGGTGGCGGCGCAAGCGCTCAGCTA 1652
DB 10852 TATGCTTGTGAACAGCCGCGAGCGCTGGCGGTAAAGCACTATGCTTATCAGTGAATTA 10911
QY 1653 CGAAGAGCTTTGCGCGCTTGGCGGCACTTGGCGCGCGGCTGCGCGAGCGGGGACG 1712
DB 10912 TCATAGAGTGAACAGCTATGCGCGCGCGCTGGCGGGGAGTTAATCAGTGGGGGTTCA 10971
QY 1713 CCGCAACATTTGGTTCGCGGTGTGTATGAGAAAGGCTGGAGAGAGTGTTCGCGCTTCT 1772
DB 10972 GCCCGCATATATGTGTATCAAGATGTCAAAAGGCGCAGAGCAAACTTGTGTGCGTTCT 11031
QY 1773 GCGGCTGTCAAGTCAAGCGCGGCTTACGTGCTGCGATCGATCGGCGGCGGACG 1832
DB 11032 GCGCTGTCTGTGGCGGGGCGGTTTACGTTTCGCGTTTCTGCTGATCAGCTTCCGCGACG 11091
QY 1833 TATCCACTACTCTCTCATCATGTGTGAGTTAAAGCTGCTGTGACGAGCATGGCTGGA 1892

```

Db 11092 GGGCGAGAAATCTACGCTGAGCCGCGCTCGCTGCTCATTTTGTACAGACAGCC 11151
Qy 1893 TGGCAAACTGTATGCGCGCGCGGATCCAGCGGCTGCTGTGAGCGAGCCGGCTGCA 1952
Db 11152 CAGCGCGGGGTCA-----GACGATATTCGCGCTTGCCTGGCAGAGCCATT----- 11200
Qy 1955 AGGCGACGGCGCAAGCCTCCGATGATGCGCATTCAGACACTTGGATCTGCGGTATG 2012
Db 11201 ---GAGCGGAGCGCGATGCGCAACCGGATGAGCCGCGCCCGACGCAACCGGCTTACAT 11256
Qy 2013 CATCTACACCTCGGGATCCAGCAAGGTTGCGCAAGGGGGTGTATGATCATCGAGGTG 2072
Db 11257 TATCTACACCTCGGGTCTACCGGATACGCGAAGGGGATGATTTCTCACCGGGAGAC 11316
Qy 2073 CGTCAACACCATCTGGAATCAACGAGCGCTTCGAATAAGGCGCGGAGACAGGTTGCT 2132
Db 11317 GCTTAAACCTGTGCGATATCAATACCGCTATCAGTTGGCGCCGATGACAGGGTGTCT 11376
Qy 2133 GGGCGCTCTCGCTGAGCTTGTGATCTCGGCTATGATGTTTGGGATCTGGCGGG 2192
Db 11377 GGGCGCTCTCGGCTTACATTTGATTTATCGGTTTACGACATTTTGGCGTACTGGCGC 11436
Qy 2193 GGGCGGTAAGATCGTGTGCTCCGAGCGCTTCAAGCTGCGCGATCCGCGCATTTGGGAGA 2252
Db 11437 GGGCGCGCGCTGTGTATGTGTATGTAATAATCAAGCGCGCATCTCAGCATGTGTGTA 11496
Qy 2253 GTTGTATGAGAGAGAGAGTGAAGGTGTGTAATCTCGTGTGCGGGCGCTGATGCGAGTCT 2312
Db 11497 GCTGATCCAGCGCATCAAGTCAAGTCAAGTCAAGCGCGCGCTGCTGATGATGCT 11556
Qy 2313 CGTCAAGCATTTTGGGGGTGCGCGCGGATTCGCTGCTAGGTCTGCGGCTTTGCTGCT 2372
Db 11557 GCTGACCTGTGTGAAGGTTTGGCGGACGCGACGCGGAAACCTGCGCGCATGTATGCT 11616
Qy 2373 GAGCGCGCATGATGATCCCGGTGCGCTGCTGCGAGCTCAAGCCATCAAGCGCGCGCT 2432
Db 11617 TTCGCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11676
Qy 2433 GTGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2492
Db 11677 ACAATTTATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11736
Qy 2493 GAGGAACCTGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2552
Db 11737 TCACGACGTCGCCCGCGCTGCGCTGCAATCCTTACGCTTTCGCTTAAACCAACGAG 11796
Qy 2553 GTTCCACGTCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2612
Db 11797 CTAACGCGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11856
Qy 2613 CATTTGGCGGGTCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 2672
Db 11857 GATTTGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 11916
Qy 2673 CTTTCCGTCGCAACCGCGAGACCGGCGCTTCAAGACCGGAGATCTGGGCGGCTA 2732
Db 11917 ATTTTTCGCTCCGGA-----CGAGGCTGTGTATGCGACCGGAGATCTGGGCTGCTA 11970
Qy 2733 CTTGCGCGTGAACCGCGAGACCGGCGCTTCAAGACCGGAGATCTGGGCGGCTA 2792
Db 11971 CTTGCGCGTGAACCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12030
Qy 2793 ATACCGCGTGAACCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2852
Db 12031 ATATCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 12090
Qy 2853 CGGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2873
Db 12091 AGCAACCGTTCGCGCATGCG 12111

```

RESULT 8
ABX34289

```

ID ABX34289 standard; DNA; 135638 BP.
XX AC ABX34289;
XX 11-FEB-2003 (first entry)
DT S. atroolivaceus leinamycin biosynthesis gene cluster.
XX DE
XX S. atroolivaceus leinamycin biosynthesis gene cluster.
XX leinamycin biosynthesis gene cluster; lnm; open reading frame; ORF;
XX anti-tumour antibiotic; broad spectrum antimicrobial activity;
XX Gram-positive; Gram-negative bactericidal; chemical modification;
XX metabolite; apo-carrier protein; holo-carrier protein; tumour;
XX polypeptide; hybrid polypeptide/polypeptide metabolite; lnm production;
XX cytosolic; gene; ds.
XX OS
XX Streptomyces atroolivaceus.
XX PN
XX MO200271179-A2.
XX PD
XX 03-OCT-2002.
XX PF
XX 22-MAR-2002; 2002WO-US08937.
XX PR
XX 26-MAR-2001; 2001US-278935P.
XX PA (REGC ) UNIT CALIFORNIA.
XX PA (KYO) KIOMA HAKKO KOGYO KK.
XX PI
XX Shen B, Cheng Y, Tang G;
XX WI: 2003-018907/01.
XX P-PDB: ABU11341, ABU11342, ABU11343, ABU11344, ABU11345, ABU11346,
XX ABU11347, ABU11348, ABU11349, ABU11350, ABU11351, ABU11352, ABU11353,
XX ABU11354, ABU11355, ABU11356, ABU11357, ABU11358, ABU11359, ABU11360,
XX ABU11361, ABU11362, ABU11363, ABU11364, ABU11365, ABU11366, ABU11367,
XX ABU11368, ABU11369, ABU11370, ABU11371, ABU11372, ABU11373, ABU11374,
XX ABU11375, ABU11376, ABU11377, ABU11378, ABU11379, ABU11380, ABU11381,
XX ABU11382, ABU11383, ABU11384, ABU11385, ABU11386, ABU11387, ABU11388,
XX ABU11389, ABU11390, ABU11391, ABU11392, ABU11393, ABU11394, ABU11395,
XX ABU11396, ABU11397, ABU11398, ABU11399, ABU11400, ABU11401, ABU11402,
XX ABU11403, ABU11404, ABU11405, ABU11406, ABU11407, ABU11408, ABU11409,
XX ABU11410, ABU11411.
XX PT Novel gene cluster responsible for synthesis of leinamycin in
XX Streptomyces atroolivaceus useful for making various peptide and/or
XX polypeptide, and/or hybrid polypeptide/polypeptide metabolites -
XX PS
XX Claim 6; Page 81-127; 185pp; English.
XX The present invention relates to the isolation of the Streptomyces
XX atroolivaceus leinamycin (lnm) biosynthesis gene cluster containing
XX 71 open reading frames (ORFs) (ORFs -35 through -1, ORFs lnm through
XX lnmf, and ORFs +1 through +9). Leinamycin is a novel anti-tumour
XX antibiotic produced by several Streptomyces species. It exhibits
XX broad spectrum antimicrobial activity against Gram-positive and
XX Gram-negative bacteria, but not against fungi. The polypeptides encoded
XX by the lnm biosynthesis gene cluster ORFs are useful for chemically
XX modifying a molecule in a host cell. The host cell is a bacterium or
XX eukaryotic cell, including a mammalian, yeast, plant, fungal, or insect
XX cell. The molecule is an endogenous metabolite produced by the host
XX cell or exogenously supplied metabolite, or an amino acid, and the
XX polypeptide is a peptide synthetase or amino transferase. The
XX polypeptides encoded by the lnm gene cluster are useful for converting
XX an apo-carrier protein to a holo-carrier protein. lnm shows potent
XX antitumour activity in tumour models in vivo. The lnm gene cluster
XX modules and/or catalytic domains are useful for making various peptide
XX and/or polypeptide, and/or hybrid polypeptide/polypeptide metabolites.
XX The proteins encoded by the ORFs are useful alone, or in combination
XX with other active domains to modify various target substrates. The
XX lnm gene cluster is useful to upregulate endogenous lnm production to
XX permit lnm production in cells and/or to make various modified lnm.
XX lnm, its analogue, or other polypeptide, peptide or hybrid
XX polypeptide/peptide metabolites are useful as therapeutic agents, to

```


Db 2671686 CGACGCGCGGTGCTGCTGTAAGTCCGCGGAGGGGGTCTTTGACACCGCGCTCAT 2671627
Qy 1425 GGAACGATGCTGGAAGCGTACGTTTCTCGGCGGCTCACT-----GAGAAC 1478
Db 2671626 CGACGCGAATGTTACCCACCAAGGTGACGAATGCTCCGGTTGGCCGCCGGGAGACACGC 2671567
Qy 1479 ATGGGGTGAACAGGTGCGCTGTTGCTTCCGCTGCGCAGTAAAGCGGGCGAGCGC 1538
Db 2671566 CTGGAGTGGCGCGAGCCCGCTCGCGCTACCGCGCGAAGCGCGGTGGCGCGCT 2671507
Qy 1539 AAACGCAACAAAGCGCTGCTGAGCAGATAGCTGCAAGCGCTGTTGCGCGCGCGGT 1598
Db 2671506 GAAAGGTGCAACCGCGCCCGCCAGCAGACGAGGGGCTGCAAGAGGTTTTTCCGCAAGC 2671447
Qy 1599 CGAGAGCTGCCCAATGCACTGCGCGCTGGTGTGCGCGCGCAAGACGCTCACTGAGAA 1658
Db 2671446 CCAACGACGACCGGACGCGCGCGGTGTCGCAATTCGCGGCACTTACGACCCCA 2671387
Qy 1659 GCTTTCGCGCGCTGCGCGGCACTTGGCGCGCGCGCTGCGCGAGCAGGGGGCAAGCCGAA 1718
Db 2671386 ACTGCGCGACAGGCAATCGAGCGGTGCGCGCGCTGCTGCGCGCGCTACGAGTCCG 2671327
Qy 1719 CACATTTGCTGCGGTGATGAGAAAGCTGGAGCAGATTGTCGCGTTCTCGCGT 1778
Db 2671326 CGACACCGTCCGGGTGCTGGGTCCGAAAAGCGGCAAGAGTGGCGGTGTGCTGGGAT 2671267
Qy 1779 GCTCGAGTCAAGGCGCGGCTTACGCTGCGCATTCGATCCGACCTAACCGCGAGGCTATCA 1838
Db 2671266 TTTGGCGCGCGGGGTCTATCTGCGCATCGCGCTGCAAGCCCGCGACCGCGCGA 2671207
Qy 1839 CTACCTCTCGATCATGTAGGTAAAGCTCGTGTGAGCAGCAGCCATGCGGATGGCA 1898
Db 2671206 GCGCATCTCGGCGACCGGTTCCGTCACTT-----AGCGTCTGTTGCGG 2671162
Qy 1899 ACTGTCATGCGCGCGCGGATCCAGCGGCTGCTGTGAGCAGGCGCGCTCGAAGCGA 1958
Db 2671161 CCGCGCATGCTCAAGTGGGATGCGGTCCGAGACGCTGTTGCTGAGGTGTTGCCG 2671102
Qy 1959 CGGCGACACGCTCCGATGATGCGCATTCAGACACTTGGATCTGCGTATGCTATCTA 2018
Db 2671101 CCGCGCGGAGAATGTCGCCCGGCTAGCGATCCACCGCGCTGCTATGTTGTT 2671042
Qy 2019 CACCTGGGATCCAGAGGGTGGCGCAAGGGGGATGATGATCATCGGGGTGCCGTCA 2078
Db 2671041 CACCTGGGCTCAACCGGGAAACCAAGGGTGTGAGTGGCGCAAGACCGCGATGA 2670982
Qy 2079 CACCATCTGACATCAACGAGCGCTTCGAAATAGGCGCGGAGACAGGGTGTGCGCT 2138
Db 2670981 CACCGTGAAGACCTTTCATCGGCACTTCGAGCTAGGCGCGGAGATCGGTGCTGCCGT 2670922
Qy 2139 CTGCTGCTGAGCTTGAATCTCTCGGTCTATGATGTGTTGGGATCTTGGCGGGCGG 2198
Db 2670921 GCGGACGCTGAGTGGAGATGTCGTTGTCGATCTTCGCGCGCTCGCTCGCGCG 2670862
Qy 2199 TAGATCGTGGTGGCGGAGCGCGTCCAAAGCTGCGGATTCGCGCGCATTTGGGAGAGTTGAT 2258
Db 2670861 AGCGATGCTGGTGTGAGAGAACGCAAGCGCGGATTCGCGAGCCTGGGCGCGCTTAT 2670802
Qy 2259 CGAACGAGAAAGTGAAGTGTGAATCTCGGTGCGCGCGCTGATGCGATGCTCTGCA 2318
Db 2670801 CGACACTTACGAGGTCAACCGCGTGAATTCATGCGGGGTGGTTGACATGCTGCTCA 2670742
Qy 2319 GCATTTTGAAGGTGCGCCGATTCGCTCGTAAAGTCTTGGCGCTTTCGCTGTAGCGG 2378
Db 2670741 AGTGGCGGGGGCGCGCTG-----TCGTGCTGCGAGCGGTGCGCTGCGCG 2670694
Qy 2379 CGACGTGATCCCGGTGGGCTGCTGCTGAGCTCAAGGCGCATAGGCGCGCGGTGCGGT 2438
Db 2670693 CGACTGGGTGCTGCTGCGACCTGCGCGCGCTGCAAGTGGCAAGCCCGAGCCACGTTT 2670634
Qy 2439 GATCACTGGCGGGGCGGACCGAAGCTGATCTGCTCATCGGGTACCCCTGAGAA 2498
Db 2670633 CGGGGGGTTGGGTGAGCCACGAAACCGGTGTCACGACCATTTTGAAGTCCAGGA 2670574

Qy 2499 CGTGCAC-----CTATCGTGGCGAGCATGCCCTACGCGCGCTGCCGCAACA 2549
Db 2670573 CGCGGCAATCTGCCACGCGACTGGGCTCGGTGCAATACGCGCTCCGTTCCAAACA 2670514
Qy 2550 GACGTTTCAACGCTGCTGATGAGCGCTCGAACCAGCGCGGTCTGGGTTCCGGGCAACT 2609
Db 2670513 CGCTGCGCGGTGATGAGCGCGACAGCGGCGACGACACTGCCCATTTGGGTGGCCGATGATT 2670454
Qy 2610 CTACATGGCGGGGTGCGGCTGCGCACTGGGCTACTGCGCGCATGAAGAGACCGCA 2669
Db 2670453 GTGGGTGTCGCGTCCGGAATCGCCGGGGTATCCGTGCGCTCCGAGTGAACCGCA 2670394
Qy 2670 GAGCTTCTCTGTCGACCCCGAGACCGGGGAGCGCTTCAACAAGCGCGGATCTGGGCG 2729
Db 2670393 GCGCTTCTGTCGAGTACGCGCGCA-----CTGTATGCAACCGGTGATCTGGCCG 2670340
Qy 2730 CTACCTGCCCGATGAGAAATGAGTTGATGAGGCGGTGAGCAACCAATCAAGCTTCG 2789
Db 2670339 CTACTGGACGACGCGCACTTGAATGTTGTTGGCGGTCCGATCAACGGTCAATCAG 2670280
Qy 2790 CGGATACCGCGCTTGAAGCTGCGGGAATGAGAGAAAGCTCAAGTCCGATCCGATCAG 2849
Db 2670279 CGGTAACGCGCTGCAACTCGCGGATGCAAGCCGCTGCAAGCCGCTTGGCGGTGCA 2670220
Qy 2850 CGACGCGGTG 2859
Db 2670219 CGCGCGCGCG 2670210

RESULT 11
AAS53964
ID AAS53964 standard; DNA; 1416 BP.
XX
AC AAS53964;
XX
XX
DT 13-FEB-2002 (first entry)
XX
DE Klebsiella pneumoniae DNA for cellular proliferation protein #93.
XX
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Klebsiella pneumoniae.
XX
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-20727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU,
PI Yamamoto RT, Xu HH;
XX
DR P-PSDB; AAU36105.
XX
XX
PT New polymucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Claim 27; Seq ID No 7601; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 1416 BP; 271 A; 425 C; 429 G; 291 T; 0 other;

CC Query Match 7.2%; Score 305.6; DB 23; Length 1416;

CC Best Local Similarity 54.4%; Pred. No. 1.7e-55; Indels 21; Gaps 3;

CC Matches 692; Conservative 0; Mismatches 559; Indels 21; Gaps 3;

QY 1602 GCAGCTGCCCATGACGCTGCGGTGTGCGCGCGCAAGACGCTACAGTACGAGACCT 1661
 DB 57 GCAACAGCGCGAGGCGCTGGCGGTAAACGACATGCGTTATCGATGGAATTATCATGAGCT 116
 QY 1662 TTCCGCGCGGTTCGCGGCACTTGGCGCGCGCTGCGCGAGCAGGCGGCAACAC 1721
 DB 117 GACACACTATGCGCCCGTTCGCGCGGACAGTTAATCGAGTCGCGGCGTTACGCCGCGCA 176
 QY 1722 ATTGTCGCGGTGTGATGAGAAAGCGTGGAGAGGTCGCGGTTCTCCGCGTCT 1781
 DB 177 TAAATGTGCTACAGATGTGAAAGCGCAGACAACTGTGTGCGTTCTGCGCGTCT 236
 QY 1782 CGAGTCAGGCGCGGCTTACGTCGCGATCGATCCGACCTACGCGGAGCGTATCAGTA 1841
 DB 237 GCTGCGCGGCGGCGTTTACGTTCCGTTTCCGTGATCAGCCTCCCGCAGCGCGGAGAA 296
 QY 1842 CCTCTCATCATGTGTGAGTAAGCTCGTGTGACGACGACCGATGCGTGAATGCAAACT 1901
 DB 297 AATCTACCTGACGCGCAGCGTCCGCGTGGTCTCA-----TTGTGACAGCAGACGCGAG 350
 QY 1902 GTTCATGCGCGCGCGGATTCAGCGGCTGTCTGTGACGAGCGCGCGCTCGAAGCGAGG 1961
 DB 351 CGCCGCGGTACAGCATATTCCTGTCCTGCGAGCAGCGCATTTGAGCGCGACCGAT 410
 QY 1962 CGACGAGCTCCGATGATGATCCCATTCAGACACTTCGATTCGCGGTATGTCATCTAC 2021
 DB 411 CGCCAAACC-----GATGATACGCGCGCCCGCAGCAGCGGCTCATATTATTTACAC 461
 QY 2022 CTCGGGATTCACAGGCTTCCCAAGGCGGTATGATCATCTCGGGGTGCGCTGAACAC 2081
 DB 462 CTCGCGCTCTACCGGTACGCGGAAAGGGGTATCATTTCTACCGGGAGCGCTTAACAC 521
 QY 2082 CATCTGACATCAACGAGCGCTTGAATAAGCGCGCGAGACAGGGTGTGGCGCTCTC 2141
 DB 522 CTGTTGCAATACAAATACCGCTATCAGGTGGCGCGCATGACAGGCGTGGCGCTCTC 581
 QY 2142 CTCGCTGAGCTTCATCTCTCGGTCTATGATGTGTTCCGGAATCTGCGCGCGGCGGTAC 2201
 DB 582 CGCCCTTACATTTTATTTATCGTTTACGACATTTTGGCGTACGCGCGCGGCGGCGAC 641
 QY 2202 GATCTGTGTGCGGAGCGGTCCAGCTGCGCGATCCGCGCATTTGGCGCAAGTTGATGCA 2261
 DB 642 GCTGTGTGTGTGATGAGAAATCAACGCGCGATCTCTACCGATGTGTGAGCTATCA 701
 QY 2262 ACGAGAGAGTGTGAGGTGTGAACTCGGTGCGCGCGCTGAGCGGAGTCTCTGTGAGCA 2321
 DB 702 GCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 761

QY 2322 TTTGAGGCTGCCCGCATTCGCTCGTAGTCTCTGCGGCTTTGCTGTCGAGCGGCA 2381
 DB 762 GTGTGAAGTTTTCGCCGACGCGCACCGCGAATCTTGGCGCAGATGATCTTTCCGCGCA 821
 QY 2382 CTGATCTCCGCTGCGCTGCTGCGAGCTCAAGCCATCAAGCGCGCGCTGTGAT 2441
 DB 822 CTGATCTCGGCTTGAACCTCCCGCGCTTATCGGCGCTTCCGCGCACAAAGCAATTTAT 881
 QY 2442 CAGCTGGCGGCGGCGCACCGAAGCGTGTGATCTGATCGGATACCCCGTAGGAGAGT 2501
 DB 882 CGCGATGCGCGCGCGCACCGAGCGGTCTATTTGCTTACCGCTCCGCAATTCACGAGT 941
 QY 2502 CGACCTATCTGTGCGAGACATCCCTTACGCGCTGCTGCGCAACAGAGCTTCCAGCT 2561
 DB 942 CCGCGCCACTGCGGCGCTCATCTCTTACGTTTCCGTACCAACCAACGCTACCGGCT 1001
 QY 2562 GCTGATGAGCGCTCGAACCGCGCGCTGCGTTGCGGCGCACTTACATTGCGCG 2621
 DB 1002 GGTGATGAAACAGGCGCGCGACTGCGCTGAGTGGGCTGATATGATTTGCGCG 1061
 QY 2622 GGTGCGGCTGCGACTGCGCTACTGCGGATGAAAGAGACGCGCAAGAGCTTCTCGT 2681
 DB 1062 CATTTGGGCTCGGAGAGCTATTTCAACGATCCCTGCTGACGACGACATTTTAC 1121
 QY 2682 GCACCCCGAGACCGGCGAGCGCTCTTACACAGCCGCGATCTGGGCGCTTACCTGCCGA 2741
 DB 1122 GCTCCCGA-----CGAGCGCTGTATCGACCGCGGATCTCGGCTCTTACGCGCAGA 1175
 QY 2742 TGGAAACATCGATTTTATGAGGCGTGAAGACAAATCAAGCTTCCGATACCGGCT 2801
 DB 1176 TGGACATCGATTTCTCGGTGTGCGGACAGAGTCAAGTGTGAGATATCGAT 1235
 QY 2802 TGAGCTGGGGAATTCAGAGAAACGCTCAAGTCCGATCGAACGTCGCGCGTAT 2861
 DB 1236 CGAGCTGGCGAATTCAGAACCGCGCTCAGCAGCTGCGGCGGTGAACAAAGCAACCT 1295
 QY 2862 TGTGCGCGTGG 2873
 DB 1296 TCTGGCATCGG 1307

RESULT 12
 AAF81367
 ID AAF81367 standard; DNA; 7335 BP.
 XX
 AC AAF81367;
 XX
 DT 04-JUN-2001 (first entry)
 XX
 DE Quorum sensing controlled gene *qscl09* ORF.
 XX
 KW Quorum sensing; antibacterial; bacterial signalling;
 KW opportunistic pathogen; immunocompromised; burn; cystic fibrosis;
 OS immunosuppressive therapy; AIDS; ss.
 XX
 OS *Pseudomonas aeruginosa*.
 XX
 PN W0200118248-A2.
 XX
 PD 15-MAR-2001.
 XX
 PF 01-SEP-2000; 2000WO-US24141.
 XX
 PR 03-SEP-1999; 99US-0153022.
 XX
 PA (IOWA) UNIV IOWA RES FOUND.
 XX
 PA (QUORUM) QUORUM SCI INC.
 XX
 PI Whiteley M, Lee KM, Greenberg EP, Muhl U;
 DR WPI; 2001-265973/27.
 XX
 PT Identifying modulators of quorum sensing signaling in *Pseudomonas*

FT	CDS	12212..13658	/*tag= e	/product= "ORF5-3-oxoacyl-ACP-reductase"	FT
FT			/note= "ACC start codon"		FT
FT	CDS	15374..19984	/*tag= f	/product= "ORF6-polyketide synthase"	FT
FT			20003..27889		FT
FT	CDS		/*tag= g	/product= "ORF7-peptide synthetase"	FT
FT	CDS	28251..29400	/tag= h		FT
FT			/product= "ORF8-transpeptidase"		FT
FT	CDS		complement (30040..31720)		FT
FT			/*tag= i	/product= "ORF9-regulation element"	FT
FT			/note= "CGC stop codon"		FT
FT	CDS	31982..32932	/*tag= j	/product= "ORF10-transcription regulator"	FT
FT			33128..33613		FT
FT	CDS		/*tag= k	/product= "ORF11-regulation element"	FT
FT			/note= "GTG start codon"		FT
FT	CDS	33661..34077	/*tag= l	/product= "ORF12-regulation element"	FT
FT			complement (35255..35616)		FT
FT	CDS		/*tag= m	/product= "ORF13-transcription regulator"	FT
FT			complement (35730..36242)		FT
FT	CDS		/*tag= n	/product= "ORF14-transcription regulator"	FT
FT			/note= "GTG start codon"		FT

XX DEL19846493-A1.
 PD 13-APR-2000.
 XX
 XX 09-OCT-1998; 98DE-1046493.
 XX
 XX 09-OCT-1998; 98DE-1046493.
 PR
 XX (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 XX
 XX Beyer S, Mueller R;
 PI
 XX WPI; 2000-294101/26.
 DR
 XX
 XX DNA sequence coding for products involved in the biosynthesis of
 PT polyketide or heteropolyketide compounds, especially epothilone -
 XX
 XX
 PS Claim 3; Page 20-33; 36pp; German.
 CC
 CC This invention describes a novel DNA sequence (I) whose expression
 CC products effect or are involved in the enzymatic biosynthesis,
 CC mutasynthesis or partial synthesis of polyketide or heteropolyketide
 CC compounds (II). (I) can be inserted into an expression vector and used
 CC to transform or transfect prokaryotic or eukaryotic cells with the aim
 CC of obtaining strains that produce large amounts of polyketide or
 CC heteropolyketide compounds, especially epothilones, which have cytotoxic
 CC and/or immunosuppressant and antibiotic and antifungal activities and
 CC are useful as plant-protection agents. This sequence represents the DNA
 CC sequence isolated from *Sorangium cellulosum* which is described in the
 CC method of the invention.

Query Match	6.3%	Score 265,	DB 21,	Length 37856,
Best Local Similarity	46.7%	Prod. No. 1.6e-46,		
Matches 1219, Conservative	0,	Mismatches 1325,	Indels 69,	Gaps 9,
Sequence 37856 BP; 5655 A; 13666 C; 12913 G; 5622 T; 0 other;				

401 TGCAGGTGATCGAGCCTAAAGTCGACGCCGACATCGAGATCATCGATCTGCCGGGCTCG 460

Db	20259	TCGAGGTGATCGCCCCCTCGCTCGCATCGAGTGCCTCCGCTCGTGACCTGAGGAGATG	20318
Oy	461	ACCGAGACACAGGAGAGCGAGGCTCGTGTCTGTTCGAGATGCGATGTCGACCCGATCT	520
Db	20319	ACGAGCCGAGACGAGAGCGGAGCCCGCGCTCATGCGAGAGAGACCCGCGCCCT	20378
Oy	521	ATGACACCGAGCGCCCTCGCTCTATACGTCGTGCGCGTGGCTGACGAGCGGCAAA	580
Db	20379	TCGATCTCAGCGAGAGGCGCTCTCCGAGCCAAAGCTGCTCCGCTCGCGAGCGGATC	20438
Oy	581	CCCGTCTCGTGCTCAGTATCGATCTCATTAAGTTACCTAAGGACACCTGTCCATCT	640
Db	20439	ACGTGCTGATCTTGACGATCATATGCTCTCCACGCGCTGTGTGATGAGAGTCTGT	20498
Oy	641	TCAAGA-----CTGCGTCAAGCTTTCACAGAAATCCCGAGACTCTCTTCCTTG	688
Db	20499	TCAAGAGCTTTTCCAGCTCTACGCGCGCTTCCACGAGGCGCGCCGTGCGCGCTCCG	20558
Oy	689	TCTCGAGCTCTCGTACCGCGATTAATGTACTCGCGCTGAGTCTCGCAAGAAATCTAG	748
Db	20559	AGCTGCGGATTCATATACGCGGACTTTCGCGGTGTCAGCGGAGCTGTCTCAGGCGAAG	20618
Oy	749	CGATCAACGATGAGATGATTTACTTGGAGCGGCGCATCCCGAGCTCCCACTCCGCGCA	808
Db	20619	TTCTGGAAATGCACTTCGAGTACTGGAGAGACCTTCGCGCGCGCCCTCAGCTGTCTG	20678
Oy	809	CGCTTCGATGAAAGCGCGATCCATCTACCTGAAAGAGATTCGCTTCGCGCAACGAGAC	868
Db	20679	AGCTTCGAGTGAACGCGCGCGCGCGCGCGCGCGAGACGTTCCGCGGCTCCCAAGCGCT	20738
Oy	869	AATGCGTCCGCTCGGAGCTCTCTGGGGTGCATTGAAGCGCGGTGTCCGAGAGCGCGGCTGA	928
Db	20739	TCCGACTCCCACTCTCCCTGCAACAGGCGGTGAGCGGCTCAGCCGGCAGAAAGGCGCA	20798
Oy	929	CCCCGACGGGCGTCACTCTGAGCTGATTTTCCGAGTGAATCGGCGCTGAGACGCGAGCC	988
Db	20799	CCCCCTTCATGACGCTGCTGACCGCGCTTCAAGCTGTGCTCTTCGCTTTATGCGCGGACA	20858
Oy	989	CCCGGTTTACGCTCAACATTAAGGCTCTTCAACCGGCTCCCGGCTCATCCGCGGTGAAG	1048
Db	20859	GCGATCTGTGTGTGGGACCGCCATCGGATGCAACCCGAGCAGAGC-----TGGAG	20912
Oy	1049	ATATCACCGGAGCTTCACTCGATGATGCTCTCTGACATTCGACACCACTTCGCGACAGA	1108
Db	20913	GGCTGATCGGCTTCTTGTGCAACATGCTGGCGGTGGCATTCGACCTCGGGGCGACCGCA	20972
Oy	1109	GCTTCGAACAGCGCGCTAAGCTTATTCAAAGACGCTGTGGAGAGCGATGATCACTGCG	1168
Db	20973	GCTTCCCGAGCTGCTCGGCGGCTGTCGAGAGTGAAGTTGGGCGGCTTACGCGCACAG	21032
Oy	1169	ACGTAAGCGGATTCGAGGTCGACGAGAGGCGCGCGGCTCCGCGGAGTCAACGAGGCG	1228
Db	21033	ACCTGCGCTTTCGAACGCGCTGTGAGAGGTGTCACACGAGGAGAGCCCAAGCACAGCC	21092
Oy	1229	CATTGTTCCCGTGTGTCTCAGAGCGCGCTTAACGACGAATGCG-----TGTGTGTA	1282
Db	21093	CTTGTGTCAGAGTGTCTTCACTGTCAGAACACCCGATGATGAGAGAGAACAGAGCG	21152
Oy	1283	CTTCGTTGCAAGGCTTCGGAATTCGCGGTGTAACACGACGACGACTCTCAAGCTGTGC	1342
Db	21153	ACATTGATCGGGTGGCGCGCGCTGTGAATAAGAGCGGGAATTCGATCTGATCC	21212
Oy	1343	TGATGATCAGCTCTACGAGCAAGATGGGGAAGCTGTCTTCGCGTGGGACATGCG---TGG	1399
Db	21213	TGAGGCTCTGGAATGCGCGCAAGGCTTCTCGGCACTTTGAGTACAAACACCGACTGT	21272
Oy	1400	ACGAGATGTTCCGCGCGACCTTCTGAGACACATGCTCAAGCGTACGCTTTTCTTC	1459
Db	21273	TGACGCGCGGACCATTCGAGCGGATGGCGCGCACTGAGAGGTCTGTCTCCAGCGCG	21332
Oy	1460	GCGGCTCACTGAGAACCATGGGGTGAACAGGTGCGCTGTTCGCTTCGCGCTGCCAGC	1519

Db 21333 TCGCGCGCGCGGATCGACCCATTGCGAGCTGCCCTCATGCGGCGCGGAGAGCGGAGTC 21392
Qy 1520 TAGAAGCGCGCGGAGCGCAAAAGCGACCAAGCGCTGTGAGCGACATACGCTGCAG 1579
Db 21393 GGGTATGTTGGAGTGGAACTCACTGACGCGCGCTGTATCCCGAGACCAATTGATCAG 21452
Qy 1580 GCTGTGTCGGCGCGCGGCTGAGCAAGCTCCCATGAGCTCCCGGCTGTGCGCGCGCA 1639
Db 21453 ACCTGTTTCAGACGACAGTGGAGCGGTGCGCCGAGCGACCGCGGTCTCTCCACGAC 21512
Qy 1640 AGACGCTCAGTACGAGAGCTTTCCGCGCTTCGCGGCACTTGGCGCGCGCTGCGCG 1699
Db 21513 AAGAGTTGACGTATCGAGAGCTGAAATGCGCGCAATCAGCTCGGCGATCACTGCGGA 21572
Qy 1700 AGCAGGGGCGACCGCCGAAACATTTGCTGCGGTGTGATGAGAAAGCTGGAGACAG 1759
Db 21573 GCTGTGGCGTGGGCGCAGAGTGCCTGCGGTGTGATCTCGAACGCTCAATGAGACGG 21632
Qy 1760 TTGTGCGGTTCTCCGCGGTCTCGAGTCAGCGCGCGCTCAAGTGGCGGATGCGGAC 1819
Db 21633 TGTGCGGATCTCTCGGCGGTCTCAAGGCTGCGCGGCTTACGTCGCTGATCCGAGT 21692
Qy 1820 TACCGCGGAGCGTATCTCACTCTCTGATCATGATGAGTAAAGCTGCTGACCG 1879
Db 21693 ACCCGAGGAGCGCTCGGGCTCATGATGGCGGACGACGCGCTGCTGCTCAGCG 21752
Qy 1880 AGCCATGCTGATGAGCAAACTGTCAATGCGCGCGGAGATCCAGCGGCTGCTGTAGCG 1939
Db 21753 AGCGCTGCTCTCTCGAAGCTGCGCGCCGACGCGGATGCAAGCTGTGATCAGCTGAG 21812
Qy 1940 AGCGCGCGTTCGAAAGCGAGCGGAC---CAGCTCGATGATGCCATTGACAGACCTT 1996
Db 21813 CGCTGACAGAGCGCTCTCAGGCTGCCACATACCCCGAGCGGCGTACCGCC 21872
Qy 1997 CGGATCTCGGATATCATCTACCTCGGATTCACAGGTTGCGCAAGGCGGTATGA 2056
Db 21873 AGAACCTGCGATACGTATGATACCTTCGGGCTGACCGGCGGCGCCAGAGGCGTCTG 21932
Qy 2057 TCGATATGCGGCTGCGCTCAACCATCTGAGATCAACGAGCGCTTGCAGATGCGC 2116
Db 21933 TCGAGCACCGCGGCTCTGCAACCTGCCACCGTGAAGCGCAAGCTTATGAGATGCC 21992
Qy 2117 CGGAGACAGGCGTGGGCTCTCTCGTGAAGCTTCATCTCGGTATGATGAT 2176
Db 21993 CGGCGACAGGCTCTCAGTTCGCGCGCTGCTGTCTGACACATGTTCTGCGAATG 22052
Qy 2177 TCGGATCTCGGCGCGCGCGGTACGATGCTGCTGCGGACGCGTCAAGCTCGCGATC 2236
Db 22053 CGCTGCGTTGCTCTCGGAGGAGCGCTGTGATGGACAGCGGAGCGAGCTTCTCCGG 22112
Qy 2237 CGGCGATTGGCGAGGTTGATGAAAGAGAGAGGATGACGCTGTGAACTCGTCCGG 2296
Db 22113 GACCTCGGCTGCTGAGCTGCTGAAAGACACCGGCTCAAGGAGTCTCTGCGCCCTA 22172
Qy 2297 CGGTATGCGGATGCTGCTGAGCATTTTGAAGGCTGCGCCGATTCGCTGAGGCTC 2356
Db 22173 CGGTGCTCGAGCGCTGCGCAAGACAGAGCGC-----GCGTTCGCGC 22217
Qy 2357 TGGGCTTTCGCTGAGCGCGGATGATCCCGTGGGCGCTGCTGCGAGCTCCAG 2416
Db 22218 TGGCGCTGTCAAGATGCGCGGTGAGCGGTGCCCGCGGAGCT-----GTCAGCGCT 22271
Qy 2417 CCATCAGGCGCGCGCTGTGATCAGCTGGGCGGCGCAACGAGGCTGATCTGAT 2476
Db 22272 GGAAGGACCGCGAGCGCGCTGTTAACTCTATGCGCGCGAGACGACCATTTGG 22331
Qy 2477 CCATCGGATACCGCGTGAAGAACTGACCTATCGTGGGCGAGATCCCTTACGCGCT 2536
Db 22332 CAGGCTCGGAGGAGCTGTCCAGAGAAC-----GATTCGCGCCATATGCGCGCT 22382
Qy 2537 CGCTGCGCAACGAGCGTTCAGAGTCTGATGAGCGCTGCAACCGCGCGGCTGAG 2596
Db 22383 CGATTGCCAATAGCAAAATCTAGTCTGAGAGAGCGCTGAGCGCGGTCCCATGCGG 22442

Qy 2597 TTCGGGCGCACTTACATTTGGCGGGGCTCGGCTGGCACTGGGCTTACTGCGCGATGAA 2656
Db 22443 TCCCGGCGAGATCTTCAATCGCGCGGCGGTGCGCCCGGAGATACAGGCGCTCGG 22502
Qy 2657 AGAAGCGCGAAGAGCTTCTCGTGCACCC-----GAGACCGGAGCGGCTCT 2707
Db 22503 ACCTACCGCGAGGATTTCTGATCCGACCTTGGGCGAACAAGGCGGCGCTCT 22562
Qy 2708 ACAAGACCGCGATCTGGGCGCTTACCTGCGCGATGAAACATCAGTTTATGGGCGT 2767
Db 22563 ATCGGACCGCGATCGGCGCGCTGCTGCGGACGAAACCTCAGTTTCTCGGTGCA 22622
Qy 2766 AGGACACCAATCAAGCTTCCGCGATACCGGCTTACCTCGGAGAAATCGAGAAAGC 2827
Db 22623 AGGAGAGAGGTGAGGCTCGGCTGTCTGATGAGCTGAGAGATCGCGGCGT 22662
Qy 2828 TCAAGTCCATCCGAACGTACGCGACGCGGTATGTCCTCGGAAACGACGCGCA 2887
Db 22683 TCTCAAGCACCGCGCGGTCTGCTCAAGCGCGTGTGCTGCGGAGACAGCGCGG 22742
Qy 2888 ACAAGCTCTTCTAGCTTATGTTGTCGCGAGGCAACGAGACGCGCTGCGAGCAG 2947
Db 22743 ACAAGCGGCTCGTCCGATATGCTGCGACGCGGAGAGCGCGCTGACCGCGGAGC 22802
Qy 2948 AGCGAGCTCAAGACCGAGCGGATCGACGCA 2980
Db 22803 TCGCGAGTCTCGTGAAGCGAGGATTGCTGCGA 22835

RESULT 14
AAS1470
ID AAS1470 standard, DNA; 13029 BP.
AC AAS1470;
XX
DT 13-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #5.
XX
KW antisense; ds; prokaryotic cellular proliferation gene;
XX antibiotic; antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
XX Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
DR P-PSDB; AUI3611.
XX
PT New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Claim 27; Seq ID No 4052; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 13029 BP; 1978 A; 4677 C; 4485 G; 1889 T; 0 other;

Query Match 6.2%; Score 261.4; DB 23; Length 13029;

Best Local Similarity 51.8%; Pred. No. 7.5e-46;

Matches 704; Conservative 0; Mismatches 631; Indels 24; Gaps 4;

QY 1572 GCTGACGAGCTGTTCCGCGCGCGGGTGCAGACGCTGCCATGACGCTGCGTGTGTC 1631
 DB 3387 GCTGCGGAGCTGCTGAGCGGCACTGCGCGAGTCCCGCAACCGCTGCGCTGAGTG 3446
 QY 1632 GCGCGCAAGACGCTACGATGAGAGCTTTCCGCGCGTTCGCGCGACTTGGCGCG 1691
 DB 3447 GAGCGCGGAGCTTCGCGTACGCGAAGTCAAGCCCGCGGCAACCGCTGCGCGCACTA 3506
 QY 1692 GCTGCGGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1751
 DB 3507 CTTGCGGCAAGAGCG 3566
 QY 1752 GAGAGAGGTTGTCGCGGTTCTCGCGGTGCTGAGTCAAGCGCGCGCGCGCGCGCG 1811
 DB 3567 GCAATGCTGCTGCGGTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3626
 QY 1812 TGCCGACCTACCGCGCGGAGCGGATGCACTCTCTGATGATGATGATGATGATGAT 1871
 DB 3627 CCCGAGCTACCCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3686
 QY 1872 GCTGACGAGCG 1931
 DB 3687 ACTGACCGAGCG 3746
 QY 1932 CGTGAGCGAGCG 1991
 DB 3747 CTTGAGCGAGCG 3806
 QY 1992 ACCTTGCGATCTCGCGTATGTCATCTACACTCGGAGATCACAAGGATGCGCGAGG 2051
 DB 3807 CGACA---ATCTCGGCTACGTCATCTACACTCGGAGTCCACCGCGCGAGGCGGT 3863
 QY 2052 GATGATCGATCATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2111
 DB 3864 GGGCAACACCCACG 3923
 QY 2112 AGGCGCGGAGACAGAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2171
 DB 3924 CGACGCGGACGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3983
 QY 2172 TGTGTTGCGGATCTGCG 2231
 DB 3984 GTGCTTCTGCG 4043
 QY 2232 CGATCCG 2291
 DB 4044 CGATCCG 4103

QY 2292 GCGCGCGCTGATGCGGATGCTGTCGACATTTTGAGGAGTCCCGCGATGCTGCTAG 2351
 DB 4104 GCGCGCGCTGCTGCACTGTTTCATC-----GACGACCCCGCGCTGCGCGCGCG 4154
 QY 2352 GTCCTGCGGCTTTCGCTGCTGAGCGCGGACTGATGATCCCGGTGGCGCTGCGGAGCT 2411
 DB 4155 CAGCTGCGCGCGCGCTGTTCTCCGCGCGGAGCGCGCTTCCGCGGAGCTGCGCAACCGCGT 4214
 QY 2412 CCAGCGCATGACG 2471
 DB 4215 GTTGCAAGCGCTGCG 4274
 QY 2472 CTGCTCATCGGATACCCCGGTAGAGAGCGTGAACCTATGCGGAGCATCCCTACCG 2531
 DB 4275 -----CAGCTGCAACCATGCGAGTGCAGTGCAGCGCGGAGACCGCGCGCGCGCG 4328
 QY 2532 CCGTCCGCTGCGCGCAACGAGCGTTCCAGCTGCTGATGAGCGCTGCAACCGCGCGCGCG 2591
 DB 4329 CCGACCGCTGCGCGCAACGAGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4388
 QY 2592 CTGCGGTTCCGCGCGCAACCTGATGAGCGGAGTGCAGCTGCGCGCGCGCGCGCGCGCG 2651
 DB 4389 CGGCGTCCG 4448
 QY 2652 TGAGAGAGAGAGCGCGCGAGAGCTTCTCG-----TGACCCCGAGACCGGCGAGCGCGCT 2705
 DB 4449 CCGCT 4508
 QY 2706 CTACAGAGCG 2765
 DB 4509 GTACCG 4568
 QY 2766 TGAGAGAGAGAGCGCGCGAGAGCTTCTCG-----TGACCCCGAGACCGGCGAGCGCGCT 2825
 DB 4569 CCGCT 4628
 QY 2826 GCTCAAGTGCATGCGAGCGTACGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2885
 DB 4629 CCGTCCG 4688
 QY 2886 GAAAGCTCTCTTACGCTATGTCGTCGCGAGGCGAC 2924
 DB 4689 CAGCGAGTGTGTCGCTACCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4727

RESULT 15
 ABQ78872
 ID ABQ78872 standard; DNA; 90600 BP.
 XX
 AC ABQ78872;
 XX
 DT 23-OCT-2002 (first entry)
 XX
 DE S. roseosporus daptomycin biosynthetic gene cluster 90kb region.
 XX
 DE Daptomycin biosynthetic gene cluster; thioesterase; antibacterial;
 XX
 DE fungicide; virucide; antiparasitic; immunomodulator; antileptic;
 XX
 DE cytostatic; gene therapy; antitumor; immunomodulatory; siderophore;
 XX
 DE anti-cholesterolemic; agrochemical; gene; ds.
 OS Streptomyces roseosporus.
 OS
 PN WO200259322-A2.
 XX
 PD 01-AUG-2002.
 XX
 PF 17-OCT-2001; 2001WO-US32354.
 XX
 PR 17-OCT-2000; 2000US-240879P.
 XX
 PR 28-FEB-2001; 2001US-272207P.
 PR 06-AUG-2001; 2001US-310385P.
 XX
 PA (MIAO/) MIAO V P W.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:41 ; Search time 15142.8 seconds

(without alignments)
11435.853 Million cell updates/sec

Title: US-10-014-717-1_COPY_11872_16104

Perfect score: 4233

Sequence: 1 atgacgacatcagctctc.....agggcaggagcagctcraa 4233

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pac:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_srs:*

12: gb_sy:*

13: gb_vl:*

14: gb_vl:*

15: em_da:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pac:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_srs:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rtd:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_ay:*

39: em_hngo_hum:*

40: em_hngo_mus:*

41: em_hngo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4233	100.0	68750	1 AF210843	AF210843 Sorangium
2	4233	100.0	68750	1 AF193029	AF193029 Sequence
3	4233	100.0	68750	6 AR199551	AR199551 Sequence
4	4233	100.0	68750	6 AR199559	AR199559 Sequence
5	4233	100.0	68750	6 AR199567	AR199567 Sequence
6	4233	100.0	68750	6 AR201097	AR201097 Sequence
7	4233	100.0	68750	6 AR208671	AR208671 Sequence
8	4114.6	97.2	58733	1 AR217189	AR217189 Sorangium
9	4114.6	97.2	71989	6 AR172664	AR172664 Sequence
10	944.8	22.3	42603	1 AF188287	AF188287 Stigmatel
11	728	17.2	77457	1 AF210249	AF210249 Streptomy
12	634.6	15.0	189050	1 AL646066	AL646066 Ralstonia
13	609.4	14.4	40156	1 AF516145	AF516145 Lysobya m
14	600	14.2	302007	1 SCO939132	SCO939132 Streptomy
15	592.6	14.0	189050	1 AL646066	AL646066 Ralstonia
16	577.6	13.6	18023	6 AX024285	AX024285 Sequence
17	550.6	13.0	311143	1 AE016762	AE016762 Escherich
18	543.4	12.8	4305	6 AX276331	AX276331 Sequence
19	543.4	12.8	12303	1 AE009297	AE009297 Agrobacte
20	543.4	12.8	31220	1 AE008316	AE008316 Agrobacte
21	543.4	12.8	48715	6 AX276312	AX276312 Sequence
22	543.4	12.8	50538	6 AX702471	AX702471 Sequence
23	538.8	12.7	5430	1 AF184622	AF184622 Pseudomon
24	526	12.4	17071	1 AE004839	AE004839 Pseudomon
25	508.4	12.0	15023	1 AF074705	AF074705 Pseudomon
26	476.8	11.3	290850	1 SCO939127	SCO939127 Streptomy
27	454.4	10.7	4317	1 AF184621	AF184621 Pseudomon
28	406.6	9.6	309904	1 AE016865	AE016865 Pseudomon
29	395	9.3	7444	1 AE013843	AE013843 Yersinia
30	395	9.3	35806	1 AF091251	AF091251 Yersinia
31	395	9.3	119443	1 YP102KB	YP102KB Yersinia
32	395	9.3	311143	1 AE016762	AE016762 Escherich
33	395	9.3	335050	1 AL414150	AL414150 Yersinia
34	390.4	9.2	32159	6 AX702468	AX702468 Sequence
35	383.6	9.1	6468	1 YEP1RP2H	YEP1RP2H Yersinia en
36	360.6	8.5	135638	1 AF484556	AF484556 Streptomy
37	358.2	8.5	10570	1 MTCY22H8	MTCY22H8 Mycobacte
38	358.2	8.5	19782	1 AE007084	AE007084 Mycobacte
39	358.2	8.5	306550	1 BX248342	BX248342 Mycobacte
40	333.8	7.9	41156	1 MSGY222	MSGY222 Mycobacte
41	329.2	7.8	276800	1 SCO939115	SCO939115 Streptomy
42	317.4	7.5	216050	1 AL646076	AL646076 Ralstonia
43	296.8	7.0	23451	1 AB070952	AB070952 Streptomy
44	296.8	7.0	302300	1 AP005034	AP005034 Streptomy
45	287.2	6.8	15719	1 AE004667	AE004667 Pseudomon

ALIGNMENTS

RESULT 1

LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000

DEFINITION Sorangium cellulosum strain So ce90 epoch1ione biosynthesis gene cluster, complete sequence.

ACCESSION AF210843

VERSION AF210843.1 GI:6724237

KEYWORDS

SOURCE

ORGANISM

Polyangium cellulosum

Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;

Sorangineae; Polyangiaceae; Polyangium.

REFERENCE

1 (bases 1 to 68750)

Mojnar,I., Schupp,T., Ono,M., Zikkle,R., Milnamow,M.,

Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,

Pred. No. is the number of results predicted by chance to have a

TITLE
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from *Sorangium cellulosum* So ce90
JOURNAL
Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE
20130945
PUBMED
10662695
REFERENCE
2 (bases 1 to 68750)
AUTHORS
Molnar, I.
TITLE
Direct Submission
JOURNAL
Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agricultures Research Institute, Inc., 3054 Cornwallie Rd, P.O. Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source
1.68750
Location/Qualifiers
/organism="Sorangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDLAFARATSARVLDHGKATRRHYLAARGIEDRLAREH
LRIQEGSPFHCMLGDLIVELLADHPYASISFPHARSLRHPDWTSDMLVDGELV
RLAARGAPGLREYEEERERARATQEARLMLAAPCPAPDLRFEDDAGLPGP
MSPEVAERRLASAYATPELCAALAMIGCAGPSPAYEMLPENLLLGFIPT
ATAAASPGTSEALRGARLFAASMEVSSKSGNIPALMERLRTIVRAMGAND
LSRFEARLAAEVRLRROPAPAGAGLAAGVSSSGRLGIVTDGALYSGGND
LYMFORGISPVYLAGTDFFELAPPLSQMLFVARNAGTISKVLTGSPILVARN
QARPSLVARGFMAVNOAVPDEPGA.PFVQRTIMEFHPTRCHPEAGSASF
LACDEHLWCELSAGRLMRPHRPA.PFVQRTIMEFHPTRCHPEAGSASF
MADPBRALIGVDRKGTVEPIVLAETRHPAPVASEDRDIPALTYPSLTINATHVL
SCASTVADYQOLMDRPMVMNRGLPFTMDRILTARS"
complement(1900..3171)
CDS
/note="ORF2: similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to Dp-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAAPGCGALGAEEOGQDAGAGALAPAREVMAEVA
ACQMPGAVLVARGDHDVDAVGTGLGSAEMRRDTPRIASMTKAVATAVMMLVE
EGLDLDSVDRMLPELARKVILARIDGPIDETPAREPITVADLMTFTMGGISDA
SSPIORAIDELGLVNAQVPMTPHGDEMIIRGLTPLMHOGAOMKMTYGLVQGV
VRRANDQGFADPVRERILAPLGMRODPVHPVADKLAPAGCCGFTDEQCEKTRMD
GMSAVASPAPFPGAGLAVSTVDYDLTFAKMLNGGVHGRLLSASVRENTADHL
TQAOKAASSFPQGFETGGMVGVTA.PDVAVSEVGRYGMDDGGTSMINDPGEL
IGIVMTQSGFLFSGALERFMRSVVATESA"
3415..5556
/note="ORF3: similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGILTEROVLISVLTALILVYARASGELARLROPEVIGELFG
GVYLGPSVIGALAPGFHRALFQEPAPGVVLSGTSWIGALLILMAGIEVDVGLIKREA
RGALASQALAIAPPLAAGA.FSALVDRPLPSGLFGLVLSVAVSIVAKVILERMS
RRSYAVQTLAAGVSEVAAVMVLVAMTSSSGSPALAVASLALSGFLFMVLVGR
LTHLMRWADATRVRSKQGVSLVLTFLAALITOLGHLPLGALFALLVLSNAPRT
NRPDLGVOTLVA.GFAPFVLVAGRWVDSOLRTPAAGVYALLATLTAAXVPA
LCARLGLRGSAVVAVGLNKKGRDILVAVIGGILISNEAYTMVAVVAVLTYTAA
SPALLMTEKRAPPTQESARLEREARARATPGVERLIVPVAHALGFRATDYES
IVASRKLTETVDTLVSQQA.PGPSRAGESSRLAAGLRVIGIRORREKGS
IGAILASRDHDLVIGARSPARAGMSFRLQDAIVORAESVILVVDPPAAEAS
ARRIVPIIGLEYSFAADLAAHVALMADEVLTAQTDPCA.VVMRRERSRVA
ASVVEAVFGRGLGVSRVHVGASRDESTRILARAPVDLVLGCVYDHQGLRL
YLGSTYESVVRVSRVALLVHAGTRQVR"

CDS
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MKDVGTRCAAGYIPGSGNGEPOMTSHETACLINASDRDA
OVALIVYSDDEPAGPARVTPARTTRVRRNDLTPETPRDIDYASVIESDAIVY
QHTRLDSQAENALSTLAIYDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHYDGRFRFGRRIGTTLGLLASMALAGCGSEKTVQGTFLAP
GADARVTDVDPDATTRLAVDVVHLSPRELAGSERFVVMORPSPSPRRVGVLD
YNADSRKGLAETTVPYANFELLITAEKSSPQSSAAVIGPISVC"
7610..11875
/gene="epoA"
/note="epoA"
7610..11875
/gene="epoA"
/note="epoA"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRPTRAEADPIAIVGASGRLPGVYIDLSGFMTLLGSDTV
GRVAPERDDAAMPDPDPDAGKTPYTPASLSVYACDAPAFGISPREALMDAHR
LILEVCEALENMAIABSAVGETGVGTGIPSEYELADLQATSAETIDAGGCTM
PSVGGRTSYALGNGPCVADVTASSLVVHACQSLRGESCTALAGVLSLSP
STLWLSSTRBALRGRCASAEADGREGGCAVVLKLSGARADGDIILAYRG
SAINHDGSGGLTVNGSSOEIVLKRALADGCAASVYGAHGTGTLGPIEIOA
LNAVYGRDVAATPLLIGSVKTLGHPYASGIGTLVLSVLSHOGQIPALHAQALN
PRISGDI.RLVYTRAKTPPMNTPRACVSSFGMSGNNHVVLEAPACTCPAP
RPAELIVASRTASALDQAKLRDHLSTYSQCLGDAFSLATTRKMERLVAAT
SREGRLALDAAAOQTS.PGAVRSIADSSROKLAFLFYQGAQTLGKMRGLVAVAT
REAPDLCTVRLFNQELDRLEPRLVMAEPASVDAALDQTAFTOPALFTEVLAALMRS
MGVEPELVAGHSIGELVAVAGVSLDAVFLVARGRLMOALPAGGAMVSEAPEA
DVAAVAVHAASVSIAVNAVDVYIAGAGPVHAIAMARGLARTALVSHAFS
PLMAPLEAFGRVAESVSRPSYIVNLSGKACTDVSRRGVARRARVVRADG
VKAHAAGGTFVEVGPSTLLGLVPAKMPDRPALLASSRAGREPTVLEALGIM
AVGGLVSAAG.FPSSGRRVPLPTYPWQERVWIDKADADGADGAGADEVEEG
GAVRGDRRSARLHPPEPSGRKEVNAORPFLLEIDEGVDHLVATYERAPG
LGEVIVADAGLSFNDVQALAGVPPDDLPGKPNPILGPGVAGRYAVGEGVNLQ
VGOPIYALASGAFHVTTSALVLPFOALSAIPAAAMPVAYLTAAVLDRIALQ
GGERV.IHAATGCVLAAYMOHGA.EVHATAGPEKRAYLESIGVRVYVSDSRDFV
ADVRATWGEGVDVYLNLSGELLIDKSNILRSRHPFELKRCQYADNOUGLPELR
NLSFSLVLRGEMERPARVRLLEBELGLIAGVFTPPITATLFIARVADFRMAQ
AOHLQKLVLTLDPEVOQIRIPHAGAGSTGDRDLDDLAAPARAALAEAFRTQ
VSOVTRTEIKGAEALFTRLGMDSLVMEJRNRIEASLKKLSTFLSTSPNALLA
ONLDALATSLSERVAENLRAGVQNDPVFSSAGQDWEIIAL"
7640..8920
/gene="epoA"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoA"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoA"
/note="Region: enoyl reductase"
11549..11764
/gene="epoA"
/note="Region: acyl carrier protein"
11872..16104
/gene="epoP"
11872..16104
/gene="epoP"
/note="BPOS P: contains condensation/heterocyclization,"

cytochrome adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="gi:6724259"
/translation="MTINQLNLEHOGIKLAADGERLQIAPRNALPNLARISEH
KSTILMLRPLPAESIVPAERHAPPLTDIOESYLGATFVSGHAYRED
CTDLDPRLSPAFKVVABHMDLRAHLPDMOVLEPVADIDILDRGRDSTREA
RLVLRDMSMRIDTERPPLXHYVAIDRQRIYLSIDLIVDLSGSLITKMDL
SFYDPEPSELSTRDYVLALESKRKSAHQSDYMKRIAEIAPPEPTLMKD
PSTLKEIRFRTEQMLPSDSWGLRKRRGEGILPTGVILAESEVIGRMASRFTL
NITLNLRLPVRVNDITGDTSMVLADITTRDKSPQRAKRIQOELMEMDCDS
GIEVREARVIGIORGALFVLTALNDQVGTSLQGTGPTVYSTOPDLIDH
QLYEHDDGLVLAVIDVGPDPDLDMLELVYVLRKTEPGEORGLPBAOLE
ARASANTALSHETLHGLFAARVEOLPMOLAVSARKITTYEELSRSRLCARLR
EQGRPNTLVAVMEKHEVOYAVLAVESGATVPIDADIPARIRHLDHGEKVLV
LTQPMIDGKLSWPGIQRLLVSEAGVEDGQPMMPLOTSDLAAYIYTSGLPK
GVMIDHGAVENTIIDINERFEIGPDRLALSLSLFSLVDVGIILAAGCTIVVPA
SKLRDPAHMAELIEREKTVMNSVPALMRMLVEHEGEPDLSLRSLSGMIPV
GLPELOAIRGVVISLIGATSEASISIGVPRNVDSIASIPIYGRPLRQTFHVID
EALRPRVWVPGQLYIGVGLALGYMRDEETKRSFLVHPETGERLXTGDLGYLD
GNIFWGEDNDQILKRGVRLGELIETLKSHPVNDVAVIPVANDANLKLAVVLP
EGTRRAEODASLTERIDARAHAAEMLSDGERVQFKLARERDLDEKVVLP
TGQDPREAGLDVVARSRVTFLEAPIFEVGRFLSLSEVEDGATLPFRVPSAG
STVPQTAAYVYSGRIEGVEGFYHYHFEHLKLKSHGEGHAYRONVDPEAA
FNLFPVRIDALIESLGSSEPRCLLEGAOALMRACNIGVCPQGFNEQVR
PVLDRSDYVHGMGSRVDPQFQVCTTLDQSSPRATTRGAPRGREDFADMLRD
FLRTLPEVWPTVFEVLDALPLTNSNGVDKALREKRDTSPPHSGFDALBEI
LVAVREVLGLEVGLQOSFVLDGATSIHIVMRSLQKRLREIATITELFOYNLDS
LASGLRDSRLDORPMQDRVREVRKRRRS"

gene
CDS

/gene="epob"
16251..21749
/note="Epos B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"

Query Match 100.0%; Score 4233; DB 1; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGACGATCAATCAGCTTCTGAAGAGCTCGAGCACCGGATATCAAGCTGGCGCGCAT	60
Db	11872	ATGACGATCAATCAGCTTCTGAAGAGCTCGAGCACCGGATATCAAGCTGGCGCGCAT	11931
Qy	61	GGGAGCGGCTCCAGATCAGGCCCCCAAGAACGCCGGAACCGGAACCTGCTGCTCGA	120
Db	11932	GGGAGCGGCTCCAGATCAGGCCCCCAAGAACGCCGGAACCGGAACCTGCTGCTCGA	11991
Qy	121	ATCTCCGAGCAAAAGACGATCTCGAGATGCTCGTCAAGAGATCCCGCGAGATCC	180
Db	11992	ATCTCCGAGCAAAAGACGATCTCGAGATGCTCGTCAAGAGATCCCGCGAGATCC	12051
Qy	181	ATCTGTCGCGCCCGACGCGAGCGGACGCTCGCTTCTCTCAAGACATCCAGATCC	240
Db	12052	ATCTGTCGCGCCCGACGCGAGCGGACGCTCGCTTCTCTCAAGACATCCAGATCC	12111
Qy	241	TACTGGCTGGGCGGACGAGCGGTTTACGGTCCCAAGGGATCCAGCGCTATCCGCA	300
Db	12112	TACTGGCTGGGCGGACGAGCGGTTTACGGTCCCAAGGGATCCAGCGCTATCCGCA	12171
Qy	301	TACGACTGTACGATCTCGAGCTGCGAGGCTGAGCCGCTTTCGGAAGTCTGCGC	360
Db	12172	TACGACTGTACGATCTCGAGCTGCGAGGCTGAGCCGCTTTCGGAAGTCTGCGC	12231
Qy	361	CGGACGACATGCTTTCGGGCGCACGCTGCGCGACATGATGACAGTATGAGCTTAA	420
Db	12232	CGGACGACATGCTTTCGGGCGCACGCTGCGCGACATGATGACAGTATGAGCTTAA	12291

Qy	421	GTCCAGCCCGACATCGAGATCATCATCTGCGCGGGCTCGACCGGACACACGGGAACG	480
Db	12292	GTCCAGCCCGACATCGAGATCATCATCTGCGCGGGCTCGACCGGACACACGGGAACG	12351
Qy	481	AGGCTCGTGTCTGCGAGATGCGATGTCGACCGCATCTATGACACCGAGCCCTTCG	540
Db	12352	AGGCTCGTGTCTGCGAGATGCGATGTCGACCGCATCTATGACACCGAGCCCTTCG	12411
Qy	541	CTCATACGTCGTCGCGCTGCGCTGGAAGAGCGGAACCGCTCGTCTGATGATC	600
Db	12412	CTCATACGTCGTCGCGCTGCGCTGGAAGAGCGGAACCGCTCGTCTGATGATC	12471
Qy	601	GATCTCATTAACGTTGACCTAGGAGCGCTGTCATCTTCAAGACCTGCTCAGCTTC	660
Db	12472	GATCTCATTAACGTTGACCTAGGAGCGCTGTCATCTTCAAGACCTGCTCAGCTTC	12531
Qy	661	TACGAAGTCCCGAGACCTCTCTCTGTCCTGAGCTCTGCTACCGGATTAATGATC	720
Db	12532	TACGAAGTCCCGAGACCTCTCTCTGTCCTGAGCTCTGCTACCGGATTAATGATC	12591
Qy	721	GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATGAGATGAGATTAATGAGGCG	780
Db	12592	GCGCTGAGTCTCGCAAGAGTCTGAGGCGCATGAGATGAGATTAATGAGGCG	12651
Qy	781	CGCATCCCGAGCTCCACCTCCGCGAGCGCTTCCGATGAAAGCCGATCCATACCTG	840
Db	12652	CGCATCCCGAGCTCCACCTCCGCGAGCGCTTCCGATGAAAGCCGATCCATACCTG	12711
Qy	841	AAGGAGATCCGCTTCCGCGACACGAGCAATAGCTCCGCTGGAATCTCTGGGCTGATG	900
Db	12712	AAGGAGATCCGCTTCCGCGACACGAGCAATAGCTCCGCTGGAATCTCTGGGCTGATG	12771
Qy	901	AAGCGGCTGTCCGCGAGCGCGGCTGACCCCGAGCGGCGTCACTCGGCTGCAATTTCC	960
Db	12772	AAGCGGCTGTCCGCGAGCGCGGCTGACCCCGAGCGGCGTCACTCGGCTGCAATTTCC	12831
Qy	961	GAGGTGATCCGCGCTGAGAGCGCGAGCCCGGCTTTCAGCTCAACATACGCTTTCAC	1020
Db	12832	GAGGTGATCCGCGCTGAGAGCGCGAGCCCGGCTTTCAGCTCAACATACGCTTTCAC	12891
Qy	1021	CGGCTCCCGTCAATCCGCGGCTGTAACGATATCAACCGGAGACTTCAGCTGATGCTCTC	1080
Db	12892	CGGCTCCCGTCAATCCGCGGCTGTAACGATATCAACCGGAGACTTCAGCTGATGCTCTC	12951
Qy	1081	CTGGAATCCACACCACTCCGCGAACAAGCTTTCGAAGCGCGCTAAGCGTATCAAGAG	1140
Db	12952	CTGGAATCCACACCACTCCGCGAACAAGCTTTCGAAGCGCGCTAAGCGTATCAAGAG	13011
Qy	1141	CAGCTGTGGGAAGGATGATCATCTGCGACGTAAGCGGTATCGAGTCCAGCGAGGCGC	1200
Db	13012	CAGCTGTGGGAAGGATGATCATCTGCGACGTAAGCGGTATCGAGTCCAGCGAGGCGC	13071
Qy	1201	GCCCGGCTCTGGGGAATCCAAGAGCGCATTTGTTCCCGGTGATGTCACGAGCGGCTT	1260
Db	13072	GCCCGGCTCTGGGGAATCCAAGAGCGCATTTGTTCCCGGTGATGTCACGAGCGGCTT	13131
Qy	1261	AACGAGCAATGCTGTGATGTCACCTGTTGCAAGAGGCTTCGGAATCCCGGTATCAACGAC	1320
Db	13132	AACGAGCAATGCTGTGATGTCACCTGTTGCAAGAGGCTTCGGAATCCCGGTATCAACGAC	13191
Qy	1321	ACGGAATCTCTCAGCTGCTGCTGATCATCAGCTTACAGACGAGATGGGAGCCTGCTC	1380
Db	13192	ACGGAATCTCTCAGCTGCTGCTGATCATCAGCTTACAGACGAGATGGGAGCCTGCTC	13251
Qy	1381	CTCGCTGGGACATGTCGACGAGATGTTCCCGCCGACCTTCTGGAAGACATGCTCGAA	1440
Db	13252	CTCGCTGGGACATGTCGACGAGATGTTCCCGCCGACCTTCTGGAAGACATGCTCGAA	13311
Qy	1441	GCGTACGTCGTTTTTCTCCGCGGCTCACTGAGGAACCAAGGGGTGAACAGTTCGCTGT	1500
Db	13312	GCGTACGTCGTTTTTCTCCGCGGCTCACTGAGGAACCAAGGGGTGAACAGTTCGCTGT	13371

QY 1501 TCGCTTCGCGCTGCCAGCTAGAAAGCGCGGCGAGCGCAACAGCGCTGCTG 1560
Db 13372 TCGCTTCGCGCTGCCAGCTAGAAAGCGCGGCGAGCGCAACAGCGCTGCTG 13431
QY 1561 AGGAGCATACGCTGACGCGCTGTTTCGCGCGCGGCTGAGAGAGCTGCGCATGACGCTC 1620
Db 13432 AGGAGCATACGCTGACGCGCTGTTTCGCGCGCGGCTGAGAGAGCTGCGCATGACGCTC 13491
QY 1621 GCGCTGCTGCGCGCGGCGCAAGAGCTGACGCTGAGAGAGCTTTCGCGCGCTG 1680
Db 13492 GCGCTGCTGCGCGCGGCGCAAGAGCTGACGCTGAGAGAGCTTTCGCGCGCTG 13551
QY 1681 CTTCGCGCGCGCTGCGCGAGCGAGCGGCGCGAGCGCGCAACATTCGCGCTGAGT 1740
Db 13552 CTTCGCGCGCGCTGCGCGAGCGAGCGGCGCGAGCGCGCAACATTCGCGCTGAGT 13611
QY 1741 GAGAAAGCTGAGAGAGAGCTTTCGCGCTGCTGCGGCTGCTGAGTCAAGCGCGCTGAC 1800
Db 13612 GAGAAAGCTGAGAGAGAGCTTTCGCGCTGCTGCGGCTGCTGAGTCAAGCGCGCTGAC 13671
QY 1801 GTGCCATGATGCGGACCTTACCGCGCGAGCGCTATCTACTCTCTGATCATGCTGAG 1860
Db 13672 GTGCCATGATGCGGACCTTACCGCGCGAGCGCTATCTACTCTCTGATCATGCTGAG 13731
QY 1861 GTAAAGCTGCTGCTGACGCGAGCGCTGATGAGCACTGTCATGCGCGCGCGGAGT 1920
Db 13732 GTAAAGCTGCTGCTGACGCGAGCGCTGATGAGCACTGTCATGCGCGCGCGGAGT 13791
QY 1921 CAGCGGCTGCTGCTGAGCGAGCGCGCTGAGAGCGCGAGCGCGAGCTCCGATGATG 1980
Db 13792 CAGCGGCTGCTGCTGAGCGAGCGCGCTGAGAGCGCGAGCGCGAGCTCCGATGATG 13851
QY 1981 CCCATTCAACACACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
Db 13852 CCCATTCAACACACTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 13911
QY 2041 CCCAAGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100
Db 13912 CCCAAGGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13971
QY 2101 CCGCTTGAATAAGAGCGCGGAGACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
Db 13972 CCGCTTGAATAAGAGCGCGGAGACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14031
QY 2161 TCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
Db 14032 TCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 14091
QY 2221 TCCAAAGCTGCGGATCGCGCGATGAGGAGAGTTGATGATGATGATGATGATGATGATG 2280
Db 14092 TCCAAAGCTGCGGATCGCGCGATGAGGAGAGTTGATGATGATGATGATGATGATGATG 14151
QY 2281 TGGAACTGCGGTGCGCGGCTGATGCGGATGCTGCTGAGCAATTTTGAAGGCTGCGCGAT 2340
Db 14152 TGGAACTGCGGTGCGCGGCTGATGCGGATGCTGCTGAGCAATTTTGAAGGCTGCGCGAT 14211
QY 2341 TCGCTGCTGATGCTTTCGCGGCTTTCGCTGCTGAGAGGCGAGCTGATGATGATGATGATG 2400
Db 14212 TCGCTGCTGATGCTTTCGCGGCTTTCGCTGCTGAGAGGCGAGCTGATGATGATGATGATG 14271
QY 2401 CCGTGGAGCTCCAGGCGCATCAGAGCGCGGAGTGTGCGGATGATGATGATGATGATGATG 2460
Db 14272 CCGTGGAGCTCCAGGCGCATCAGAGCGCGGAGTGTGCGGATGATGATGATGATGATGATG 14331
QY 2461 GAAAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520
Db 14332 GAAAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14391
QY 2521 ATCCCTTAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2580
Db 14392 ATCCCTTAAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14451
QY 2581 CCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640

Db 14452 CCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14511
QY 2641 TACTGCGGATGAG 2700
Db 14512 TACTGCGGATGAG 14571
QY 2701 GCGCTTCAAGAGCGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760
Db 14572 GCGCTTCAAGAGCGGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14631
QY 2761 GCGCGTGAAGAGCAACCAATCAAGCTTCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820
Db 14632 GCGCGTGAAGAGCAACCAATCAAGCTTCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14691
QY 2821 GAAAGCTCAAGTCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGAT 2880
Db 14692 GAAAGCTCAAGTCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGATCCGAT 14751
QY 2881 GCGGCGGAGCAAGCTCTTCTGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
Db 14752 GCGGCGGAGCAAGCTCTTCTGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 14811
QY 2941 GAGCAGAGAGCGAGGCTCAAG 3000
Db 14812 GAGCAGAGAGCGAGGCTCAAG 14871
QY 3001 GACGCTTGAAG 3060
Db 14872 GACGCTTGAAG 14931
QY 3061 GACCTGAG 3120
Db 14932 GACCTGAG 14991
QY 3121 GACGCTTGAAG 3180
Db 14992 GACGCTTGAAG 15051
QY 3181 GAGTTTGGCTGATTCCTGAG 3240
Db 15052 GAGTTTGGCTGATTCCTGAG 15111
QY 3241 AAATTCGCTTATCCATCCGCGGAG 3300
Db 15112 AAATTCGCTTATCCATCCGCGGAG 15171
QY 3301 TCCGCGCGGATGAG 3360
Db 15172 TCCGCGCGGATGAG 15231
QY 3361 TTGCTGAAGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3420
Db 15232 TTGCTGAAGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 15291
QY 3421 GGTTCGATGAG 3480
Db 15292 GGTTCGATGAG 15351
QY 3481 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3540
Db 15352 CTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 15411
QY 3541 CTGATGAG 3600
Db 15412 CTGATGAG 15471
QY 3601 GAAAGAGTTTCGCGGCTTCTGAG 3660
Db 15472 GAAAGAGTTTCGCGGCTTCTGAG 15531
QY 3661 GCGCGGCGGATGAG 3720

Db 15312 GCGGCGCGGATGAGACCCGCGGAGTTCCAGGTCTGTACGCTCGTCAGATTCTCACCG 15391
Qy 3721 AGCGCGGACGACGCGCGCGCGCCCTCCCGCGCGAGACGACTTCCGCGATATGCTT 3780
Db 15592 AGCGCGGACGACGCGCGCGCGCCCTCCCGCGCGAGACGACTTCCGCGATATGCTT 15651
Qy 3791 CGCGACTTCTTGAAGACCAAACTACCCGAGTACATGTGCTTCAAGTCTTCTGAGCTC 3840
Db 15652 CGCGACTTCTTGAAGACCAAACTACCCGAGTACATGTGCTTCAAGTCTTCTGAGCTC 15711
Qy 3841 GATGTGTTCCCGCTGACGCTTCCAGCGCAAGGTCGATCTGTAAGCCCTCCGCGAGCGAAG 3900
Db 15712 GATGTGTTCCCGCTGACGCTTCCAGCGCAAGGTCGATCTGTAAGCCCTCCGCGAGCGAAG 15771
Qy 3901 GATACCTGTCGCGCGCGGATTCGCGGAGCAACGCGCGCAACGCGCTTGAAGAGATC 3960
Db 15772 GATACCTGTCGCGCGCGGATTCGCGGAGCAACGCGCGCAACGCGCTTGAAGAGATC 15831
Qy 3961 CTGCTCGCGGTCGTAACGCGAGGTGCTCGGCTGAGAGTGTGCGGCTCCAGCAGAGCTTC 4020
Db 15832 CTGCTCGCGGTCGTAACGCGAGGTGCTCGGCTGAGAGTGTGCGGCTCCAGCAGAGCTTC 15891
Qy 4021 GTTCATCTTGTGTCGACATTCATTCATCTTCCGATGAGAGCCTGTGTCAGAGAGG 4080
Db 15892 GTTCATCTTGTGTCGACATTCATTCATCTTCCGATGAGAGCCTGTGTCAGAGAGG 15951
Qy 4081 CTGATAGAGAGATCGCCATTCACCGAGTTGTTCAGTACCGAGCTCGGCTCGTGGCG 4140
Db 15952 CTGATAGAGAGATCGCCATTCACCGAGTTGTTCAGTACCGAGCTCGGCTCGTGGCG 16011
Qy 4141 TCCGCTTTCGCGCGAGACTCGAGATCTAGATCAGCGGCGCAAGATGAGAGCCGAGTG 4200
Db 16012 TCCGCTTTCGCGCGAGACTCGAGATCTAGATCAGCGGCGCAAGATGAGAGCCGAGTG 16071
Qy 4201 GAGGTTCCGCGCGAGAGCGGCGAGAGAGTACTA 4233
Db 16072 GAGGTTCCGCGCGAGAGCGGCGAGAGAGTACTA 16104

RESULT 2

AR193029 LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 68750)

AUTHORS Schupp,T., Ligon,J.,Madison,, Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Goriach,J.

TITLE Genes for the biosynthesis of epoehilones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES Location/Qualifiers

source 1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t

Query Match 100.0%; Score 4233; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGACGATCAATCAGCTTCTGAAGAGCTCGAGCAACGAGGTATCAAGCTGGCGCCGAT 60
Db 11872 ATGACGATCAATCAGCTTCTGAAGAGCTCGAGCAACGAGGTATCAAGCTGGCGCCGAT 11931
Qy 61 GGGGAGCGCTTCAGATACAGGCGCCCAAGAACGCGCTGAACCGGACTGCTGCTCGA 120
Db 11932 GGGGAGCGCTTCAGATACAGGCGCCCAAGAACGCGCTGAACCGGACTGCTGCTCGA 11991
Qy 121 ATTCGAGACAAAGACGATCTGAGATGTCTCGTCAAGAGACTCCCGCAGAAATCC 180

Db 11992 ATTCGAGACAAAGACGATCTGAGATGTCTCGTCAAGAGACTCCCGCAGAAATCC 12051
Qy 181 ATTCGAGCGCGCCGACCGGACCGGACCGCTCGTTCCTTCACAGACATCCAAATCC 240
Db 12052 ATTCGAGCGCGCCGACCGGACCGGACCGCTCGTTCCTTCACAGACATCCAAATCC 12111
Qy 241 TACTGCTGGGCGGACAGAGCGTTTACGGGTCCCGAGGATCCAGCGCTATCGCGA 300
Db 12112 TACTGCTGGGCGGACAGAGCGTTTACGGGTCCCGAGGATCCAGCGCTATCGCGA 12171
Qy 301 TACGACTGATCGATCTTCGAGCTGCGAGGCTGAGCCGCGCTTTGGAAGTCTGCG 360
Db 12172 TACGACTGATCGATCTTCGAGCTGCGAGGCTGAGCCGCGCTTTGGAAGTCTGCG 12231
Qy 361 CGGACGACATGCTTCCGCGCCCAACGCTGCCGACATGATGACGATGATGAGCTTAA 420
Db 12232 CGGACGACATGCTTCCGCGCCCAACGCTGCCGACATGATGACGATGATGAGCTTAA 12291
Qy 421 GTGACGCGCGACATCGAGATCATCATCTGCGCGGCTCGACCGGAGCACAAGGAGCG 480
Db 12292 GTGACGCGCGACATCGAGATCATCATCTGCGCGGCTCGACCGGAGCACAAGGAGCG 12351
Qy 481 AGGCTGTGTGTCGAGATGCGATGTGCGACCGCATCTATGACACCGAGCGCCTCG 540
Db 12352 AGGCTGTGTGTCGAGATGCGATGTGCGACCGCATCTATGACACCGAGCGCCTCG 12411
Qy 541 CTCTATCACGTCGTCGCGCTGCGCTGCGTGAACGAGGCGCAACCCGCTGTGCTCATC 600
Db 12412 CTCTATCACGTCGTCGCGCTGCGCTGCGTGAACGAGGCGCAACCCGCTGTGCTCATC 12471
Qy 601 GATCTCATTAAGTTGATGAGCGAGCTGTCATCATCTTCAAGGACTGGCTCAGCTTC 660
Db 12472 GATCTCATTAAGTTGATGAGCGAGCTGTCATCATCTTCAAGGACTGGCTCAGCTTC 12531
Qy 661 TACGAGATCCGAGACCTCTCTCTGTCGAGACTCTGTAACCGGATTAATGACTC 720
Db 12532 TACGAGATCCGAGACCTCTCTCTGTCGAGACTCTGTAACCGGATTAATGACTC 12591
Qy 721 GCGCTGAGGTCTCCGACAGAGTCTGAGCGGACATCAAGATGATGATTAATCGAAGCG 780
Db 12592 GCGCTGAGGTCTCCGACAGAGTCTGAGCGGACATCAAGATGATGATTAATCGAAGCG 12651
Qy 781 CGCATCGCGAGCTCCCACTCCGCGGAGCGCTTCGATGAAGGCGGATCTTACCTTG 840
Db 12652 CGCATCGCGAGCTCCCACTCCGCGGAGCGCTTCGATGAAGGCGGATCTTACCTTG 12711
Qy 841 AAGAGATCCGCTTCCGACACGAGAGCATGAGCTGCGTGGAGCTCTGGGCTGATTG 900
Db 12712 AAGAGATCCGCTTCCGACACGAGAGCATGAGCTGCGTGGAGCTCTGGGCTGATTG 12771
Qy 901 AAGCGGCTGTGCGGAGACCGCGGCTGACCCCGAGCGGCTGATCTGCGCATTTTC 960
Db 12772 AAGCGGCTGTGCGGAGACCGCGGCTGACCCCGAGCGGCTGATCTGCGCATTTTC 12831
Qy 961 GAGGATCGGCGGCTGAGAGCGAGCGCGGTTTACGCTCAATGAAGCTCTTCAAC 1020
Db 12832 GAGGATCGGCGGCTGAGAGCGAGCGCGGTTTACGCTCAATGAAGCTCTTCAAC 12891
Qy 1021 CGGCTCCCGCTCATTCGCGCGGTGAACGATATCAACCGGAGCTTCACTGATGCTCTC 1080
Db 12892 CGGCTCCCGCTCATTCGCGCGGTGAACGATATCAACCGGAGCTTCACTGATGCTCTC 12951
Qy 1081 CTGACATCGACACCACTCGGACCAAGAGCTTCCGAACGCGCGCTTACGCTATTCAGAG 1140
Db 12952 CTGACATCGACACCACTCGGACCAAGAGCTTCCGAACGCGCGCTTACGCTATTCAGAG 13011
Qy 1141 CAGCTGTGGAGAGAGATGATCATCTGCGAGCGTAAGCGGTATCGAGGTCCAGAGAGCC 1200
Db 13012 CAGCTGTGGAGAGAGATGATCATCTGCGAGCGTAAGCGGTATCGAGGTCCAGAGAGCC 13071
Qy 1201 GCCCGGCTCTGGGATCCACAGAGCGCATTTGTTCCCGGTGCTTCAAGAGCGCGCTT 1260

Db	13072	GC	CGGGGTCTCGGGGATTC	CAACAGAGGGCA	TTGTTC	CCGTGGTGTCTA	CAAGAGGGGCTT	13131	
Oy	1261	AAC	CAGAAAGTCGTTGGTGTCA	CTGTTGC	AAAGGCTG	GGAACTC	CGGTGTACACAC	1320	
Db	13132	AAC	CAGAAAGTCGTTGGTGTCA	CTGTTGC	AAAGGCTG	GGAACTC	CGGTGTACACAC	13191	
Oy	1321	AC	GAGACTCTCC	AGCTGCTGCTGG	AATCACTCA	CGAGACAT	GTGGGACTCTGCTC	1380	
Db	13192	AC	GAGAGACTCTCC	AGCTGCTGCTGG	AATCACTCA	CGAGACAT	GTGGGACTCTGCTC	13251	
Oy	1381	CT	CGCGTGGGACAT	CGTGTGACG	AGTGTTC	CCGCCGCA	CTTCTGACGACAT	1440	
Db	13252	CT	CGCGTGGGACAT	CGTGTGACG	AGTGTTC	CCGCCGCA	CTTCTGACGACAT	13311	
Oy	1441	GC	GTAACGTCTGTTTTTCT	CCGCGGCTCA	CTAGAGAA	CAATGG	GGTGAACAGTGCCTGT	1500	
Db	13312	GC	GTAACGTCTGTTTTTCT	CCGCGGCTCA	CTAGAGAA	CAATGG	GGTGAACAGTGCCTGT	13371	
Oy	1501	TC	CGTTCCCGCTCCG	CCCAAGTAA	GGCGGGGCGA	AGCGCA	ACCAACCGGCTGCTG	1560	
Db	13372	TC	CGTTCCCGCTCCG	CCCAAGTAA	GGCGGGGCGA	AGCGCA	ACCAACCGGCTGCTG	13431	
Oy	1561	AG	CAGAGATAAGCTGTCA	CGGCTGTTTGG	CGGCGGGTTC	GAGAGACTG	CCCATGTCAAGTCTC	1620	
Db	13432	AG	CAGAGATAAGCTGTCA	CGGCTGTTTGG	CGGCGGGTTC	GAGAGACTG	CCCATGTCAAGTCTC	13491	
Oy	1621	GC	CGGTGTGTCTGG	CGCGCAAGAC	CGCTCA	CGAAGAGCTTT	CGCGCGCTTGGCGGCA	1680	
Db	13492	GC	CGGTGTGTCTGG	CGCGCAAGAC	CGCTCA	CGAAGAGCTTT	CGCGCGCTTGGCGGCA	13551	
Oy	1681	CT	TGGCCGCGCGGCTG	CGCGCAAGAGG	GGGCAC	CGCTCCAA	CAATTTGGTCCGGGTGTGTATG	1740	
Db	13552	CT	TGGCCGCGCGGCTG	CGCGCAAGAGG	GGGCAC	CGCTCCAA	CAATTTGGTCCGGGTGTGTATG	13611	
Oy	1741	GAGA	AAAGGCTTGGGAGC	AGTGTGTGGCGGTTCT	CGCGGTGTCTG	AGTCAAGTCA	GGCGGCGGCTAC	1800	
Db	13612	GAGA	AAAGGCTTGGGAGC	AGTGTGTGGCGGTTCT	CGCGGTGTCTG	AGTCAAGTCA	GGCGGCGGCTAC	13671	
Oy	1801	GT	GC	CGATTCGATTC	CGCACTTA	CCCGGCGAG	CGTATCCATCTCTCGATCA	1860	
Db	13672	GT	GC	CGATTCGATTC	CGCACTTA	CCCGGCGAG	CGTATCCATCTCTCGATCA	13731	
Oy	1861	GT	AAAGCTCTGTGTCTGA	CGCAAGCCAT	TGGCTGGA	TGGCAA	ATTGTTCATGTGCGCGCGGGGATTC	1920	
Db	13732	GT	AAAGCTCTGTGTCTGA	CGCAAGCCAT	TGGCTGGA	TGGCAA	ATTGTTCATGTGCGCGCGGGGATTC	13791	
Oy	1921	CAG	CGGCTGCTCGTGA	GAGAGGCGCGGGT	GTGAAAGGCA	CGGCGCAC	AGCCTCCGATGTATG	1980	
Db	13792	CAG	CGGCTGCTCGTGA	GAGAGGCGCGGGT	GTGAAAGGCA	CGGCGCAC	AGCCTCCGATGTATG	13851	
Oy	1981	CC	CAATTCAGACAC	CTTCGATCTTC	CGCTATGTCA	CTTCAACCTTC	CGGATTCACAGGGTTG	2040	
Db	13852	CC	CAATTCAGACAC	CTTCGATCTTC	CGCTATGTCA	CTTCAACCTTC	CGGATTCACAGGGTTG	13911	
Oy	2041	CC	CAAGGGGGGTGAT	TGATTCGATCAT	TCGGGGGTG	CCGTCAACCA	TCTGTGACATCAACGAG	2100	
Db	13912	CC	CAAGGGGGGTGAT	TGATTCGATCAT	TCGGGGGTG	CCGTCAACCA	TCTGTGACATCAACGAG	13971	
Oy	2101	CG	TTTCGAAATAGGAG	CCCGGAGACAGGG	TGTGGCGCTCT	CTCTCGCTG	AGCTTTGATCTTC	2160	
Db	13972	CG	TTTCGAAATAGGAG	CCCGGAGACAGGG	TGTGGCGCTCT	CTCTCGCTG	AGCTTTGATCTTC	14031	
Oy	2161	TC	GGTCTATATGATGT	TTTCGGGATCTCT	GCAGCGGCGGGT	TCGATGTGTGTG	CCGACCGG	2220	
Db	14032	TC	GGTCTATATGATGT	TTTCGGGATCTCT	GCAGCGGCGGGT	TCGATGTGTGTG	CCGACCGG	14091	
Oy	2221	TC	CAAGCTGC	GCATTC	CGGCGCATTT	GGGCAAGATT	TCGAA	CGAGAGGTGACGGTG	2280
Db	14092	TC	CAAGCTGC	GCATTC	CGGCGCATTT	GGGCAAGATT	TCGAA	CGAGAGGTGACGGTG	14151
Oy	2281	TG	GAATCTCGGTG	CCGGCGCTGAT	TCCGATGCT	CGACGATTTT	TGAGGGTCCGCCCCGAT	2340	
Db	14152	TG	GAATCTCGGTG	CCGGCGCTGAT	TCCGATGCT	CGACGATTTT	TGAGGGTCCGCCCCGAT	14211	

OY	2341	TCGCTGCTAAGCTCTCTGCGGGCTTTGCGCTGAGAGGGGAGACTGATCCCGGTGGGCGCTG	2400
Db	14212	TTGGTCTGCTAAGCTCTCTGCGGGCTTTGCGCTGAGAGGGGAGACTGATCCCGGTGGGCGCTG	14271
OY	2401	CTTGAGCAGCTCCAGGCGCATCAGAGCCCGGGGTGTCCGTATCAGCTCTGGGCGGGCGACC	2460
Db	14272	CTTGAGCAGCTCCAGGCGCATCAGAGCCCGGGGTGTCCGTATCAGCTCTGGGCGGGCGACC	14331
OY	2461	GAAAGCTTCGATCTGTGTTCATTCGGGTACCCCGTGAAGAAAGTGCACCTATCTGTGGGGAGC	2520
Db	14332	GAAAGCTTCGATCTGTGTTCATTCGGGTACCCCGTGAAGAAAGTGCACCTATCTGTGGGGAGC	14391
OY	2521	ATCCCTTAAGCAGCGTCGCGTCCGCGCAACAGACTTTCACGTCGTCATAGAGCGCTCGAA	2580
Db	14392	ATCCCTTAAGCAGCGTCGCGTCCGCGCAACAGACTTTCACGTCGTCATAGAGCGCTCGAA	14451
OY	2581	CCGCGCCCGGTCTGTGGTTCGGGGCAACTTCATATGGCGGGGTGGGCTTGCGACTGGGC	2640
Db	14452	CCGCGCCCGGTCTGTGGTTCGGGGCAACTTCATATGGCGGGGTGGGCTTGCGACTGGGC	14511
OY	2641	TACTGTGGCGGATGAAGAAGAAAGGCGCAAGAGCTTCTGTGTGACCCCGAGACCGGGGAG	2700
Db	14512	TACTGTGGCGGATGAAGAAGAAAGGCGCAAGAGCTTCTGTGTGACCCCGAGACCGGGGAG	14571
OY	2701	CGCCTCTACAAGACCGGGCGATCTTGGGCGGCTACCTCCGCAATGAAACAATCAGATTCAATG	2760
Db	14572	CGCCTCTACAAGACCGGGCGATCTTGGGCGGCTACCTCCGCAATGAAACAATCAGATTCAATG	14631
OY	2761	GGGCGGTGAGGACAAACCAATCAAGCTTCCGGGATACCGCGGTGAGCTCGGGGAAATCCGAG	2820
Db	14632	GGGCGGTGAGGACAAACCAATCAAGCTTCCGGGATACCGCGGTGAGCTCGGGGAAATCCGAG	14691
OY	2821	GAAACGCTCAAGTCGCAATCCGAAGTACGCGACGCGCGTATTGTGCGCTCGGAGAACGAC	2880
Db	14692	GAAACGCTCAAGTCGCAATCCGAAGTACGCGACGCGCGTATTGTGCGCTCGGAGAACGAC	14751
OY	2881	GCGGCGAAACAAGCTCCTTCTAAGCTTATGTGTCTCCGAGGGGCAACGAGAGCGCGTGCCT	2940
Db	14752	GCGGCGAAACAAGCTCCTTCTAAGCTTATGTGTCTCCGAGGGGCAACGAGAGCGCGTGCCT	14811
OY	2941	GAGCAGGACCGCGAGCCCTCAAGACCGAGCGGATGTAGACGACAGACACGCGCGCGAAGGCG	3000
Db	14812	GAGCAGGACCGCGAGCCCTCAAGACCGAGCGGATGTAGACGACAGACACGCGCGCGAAGGCG	14871
OY	3001	GACGCGTTTGACGACGCGCGAGAGGGTGCAGTTCAACCTGCGCTTGACACAGCACTCCGAGAG	3060
Db	14872	GACGCGTTTGACGACGCGCGAGAGGGTGCAGTTCAACCTGCGCTTGACACAGCACTCCGAGAG	14931
OY	3061	GACCTGGAACGAAAGCCCTGCTGTTCATCTTGAACCGGGCAGAGATCCCGGGAGAGCGGGGCTG	3120
Db	14932	GACCTGGAACGAAAGCCCTGCTGTTCATCTTGAACCGGGCAGAGATCCCGGGAGAGCGGGGCTG	14991
OY	3121	GACGCTTACCGCGGCTGCGCGCTAGCGTACCGTCCGAAACCTTCTTGAGGCCCGCATTCGCTTGT	3180
Db	14992	GACGCTTACCGCGGCTGCGCGCTAGCGTACCGTCCGAAACCTTCTTGAGGCCCGCATTCGCTTGT	15051
OY	3181	GAGTTTGGTGGATTCTGAGCTGCTTGAAGCAGGCGGTGAGCCGACGCGCGACACTTCCC	3240
Db	15052	GAGTTTGGTGGATTCTGAGCTGCTTGAAGCAGGCGGTGAGCCGACGCGCGACACTTCCC	15111
OY	3241	AAATTCGTTATCCATCGAGGGGAGAGACGTACCCGAGTCAAACTACACGATGTGTCAAA	3300
Db	15112	AAATTCGTTATCCATCGAGGGGAGAGACGTACCCGAGTCAAACTACACGATGTGTCAAA	15171
OY	3301	TCGCGGCGCATCGAGGGCGTGAAGAGGGCTTCTATTATTACACCGCTTTCAGACACCGT	3360
Db	15172	TCGCGGCGCATCGAGGGCGTGAAGAGGGCTTCTATTATTACACCGCTTTCAGACACCGT	15231
OY	3361	TTGCTGAAGCTCTTCCGATCACCGGATTCGACCGGGAGCGCACTTTGGGCAAAACTTCGAC	3420
Db	15232	TTGCTGAAGCTCTTCCGATCACCGGATTCGACCGGGAGCGCACTTTGGGCAAAACTTCGAC	15291

3421 GTGTGATGAAGCGCGCTTCAACCTCTGTTGTTGGGAGGATGAGCCATCGAGTCG 3480
15292 GTGTTGATGAAGCGCGCTTCAACCTCTGTTGTTGGGAGGATGAGCCATCGAGTCG 15351
3481 CTGATGATGATGCTCTCCGAGAAATTTTCCCTGCTGAGGCGGATATATGGCGAGCTC 3540
15352 CTGATGATGATGCTCTCCGAGAAATTTTCCCTGCTGAGGCGGATATATGGCGAGCTC 15411
3541 CTGATGAGCAGGCGCTTCTGCAACATCGCGCTCTGTCGGTGGGGCAATTCATTTT 3600
15412 CTGATGAGCAGGCGCTTCTGCAACATCGCGCTCTGTCGGTGGGGCAATTCATTTT 15471
3601 GAACAGTTTCGGCGCGCTTCTGCAACATCGCGCTCTGTCGGTGGGGCAATTCATTTT 3660
15472 GAACAGTTTCGGCGCGCTTCTGCAACATCGCGCTCTGTCGGTGGGGCAATTCATTTT 15511
3661 GCGCGGCGGATGAGCAGCGCGGAGTTTCCAGTCTGTAGCTGTGTCAGGATTCCTCACCG 3720
15512 GCGCGGCGGATGAGCAGCGCGGAGTTTCCAGTCTGTAGCTGTGTCAGGATTCCTCACCG 15591
3721 AGGCGCGCAGCAGCGCGGCGCGCTTCCGCGCGCGAGCAGCATTGCGGATTCGCTT 3780
15592 AGGCGCGCAGCAGCGCGGCGCGCTTCCGCGCGCGAGCAGCATTGCGGATTCGCTT 15651
3781 CGCGACTTCTGAGAGCAAACTACCCGAGTACATGATGCTTCACTCTTCTGAGGCTC 3840
15652 CGCGACTTCTGAGAGCAAACTACCCGAGTACATGATGCTTCACTCTTCTGAGGCTC 15711
3841 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3900
15712 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15771
3901 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3960
15772 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15831
3961 CTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4020
15832 CTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15891
4021 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4080
15892 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 15951
4081 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4140
15952 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 16011
4141 TCCGCTTTCGCGCGAGATCTGAGAGATCTGATCAGCGCGCGAGCATGAGAGCCGAGTG 4200
16012 TCCGCTTTCGCGCGAGATCTGAGAGATCTGATCAGCGCGCGAGCATGAGAGCCGAGTG 16071
4201 GAGGTTTCGCGCGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4233
16072 GAGGTTTCGCGCGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16104

RESULT 3
ARI99551
LOCUS ARI99551 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355457.
ACCESSION ARI99551
VERSION ARI99551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
TITLE Genes for the biosynthesis of epochnilones
JOURNAL Patent: US 6355457-A 1 12-MAR-2002;
FEATURES Location/Qualifiers

source 1. 68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 100.0%; Score 4233; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGACGATCATGCTTCTGAAAGCTGAGAGCAGCAGGATATCAAGCTGCGCGCAT 60
11872 ATGACGATCATGCTTCTGAAAGCTGAGAGCAGCAGGATATCAAGCTGCGCGCAT 11931
61 GGGAGCGCTTCAGATACAGGCGCCCAAGAACCCCTGAACCGGAACTCTGCTCGA 120
11932 GGGAGCGCTTCAGATACAGGCGCCCAAGAACCCCTGAACCGGAACTCTGCTCGA 11991
121 ATTCGAGCACAAGAGCAGATCTGAGATGCTCCGTCAAGAGCTCCCGCAGAAATCC 180
11992 ATTCGAGCACAAGAGCAGATCTGAGATGCTCCGTCAAGAGCTCCCGCAGAAATCC 12051
181 ATTCGCGCGCGCGCGAGCG 240
12052 ATTCGCGCGCGCGCGAGCG 12111
241 TACTGCTGCGCGAGCAGAGCGCTTACGCTCCCGAGCGGATTCACGCTATCGCGAA 300
12112 TACTGCTGCGCGAGCAGAGCGCTTACGCTCCCGAGCGGATTCACGCTATCGCGAA 12171
301 TACGATCTGATGATCTTGAAGTCCGAGCGTGAAGCGCGCGCTTTCGGAAGAGCTGCGG 360
12172 TACGATCTGATGATCTTGAAGTCCGAGCGTGAAGCGCGCGCTTTCGGAAGAGCTGCGG 12231
361 CGGACGACATGCTTTCG 420
12232 CGGACGACATGCTTTCG 12291
421 GTGACGCGCGACATGAGATCATGATCTGCGCGCGCTGACCGGAGCAGCAGGAAAGCG 480
12292 GTGACGCGCGACATGAGATCATGATCTGCGCGCGCTGACCGGAGCAGCAGGAAAGCG 12351
481 AGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
12352 AGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12411
541 CTCTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
12412 CTCTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12471
601 GATCTCATTTAAGTTGACCTTGAAGCAGCTGTCATCATCTTCAAGAGCTGCTGCTGCTGCT 660
12472 GATCTCATTTAAGTTGACCTTGAAGCAGCTGTCATCATCTTCAAGAGCTGCTGCTGCTGCT 12531
661 TACGAGATTCGAGAGCTCTCTCTCTGCTGAGCTCTGTAACGCGATTATGATCTC 720
12532 TACGAGATTCGAGAGCTCTCTCTCTGCTGAGCTCTGTAACGCGATTATGATCTC 12591
721 GCGCTGAGATCTCCGAGAGATCTGAGCGGATCAAGATGATGATTTATCTGAGAGCGG 780
12592 GCGCTGAGATCTCCGAGAGATCTGAGCGGATCAAGATGATGATTTATCTGAGAGCGG 12651
781 GCGATGCGCGAGCTCCGACCTCGCGCGAGCTTCCGATGAGAGCGGATTCATCTACCTTG 840
12652 GCGATGCGCGAGCTCCGACCTCGCGCGAGCTTCCGATGAGAGCGGATTCATCTACCTTG 12711
841 AAGAGATTCGCTTTCGAGCAGCAGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
12712 AAGAGATTCGCTTTCGAGCAGCAGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12771
901 AAGCGGCTGCTGCGGAGCGCGGCTGACCCCGAGCGGCGGATCTGCTGCTGCTGCTGCTGCT 960
12772 AAGCGGCTGCTGCGGAGCGCGGCTGACCCCGAGCGGCGGATCTGCTGCTGCTGCTGCTGCT 12831

OY	961	AAGGTGATCGGGGCGCTGGAGGCGGAGGCCCGCGGTTTACGCTCAAAATTAACGCTCTTCAAC	1020
Db	12832	GAGGTGATCGGGCGCTGGAGGCGAGGCCCGCGGTTTACGCTCAAAATTAACGCTCTTCAAC	12891
OY	1021	CGGCTCCCGCTCAATCCGCGCGGTGAAAGGATATCAACGGGAGCTTTCAGTGCATGTCCTC	1080
Db	12892	CGGCTCCCGCTCAATCCGCGCGGTGAAAGGATATCAACGGGAGCTTTCAGTGCATGTCCTC	12951
OY	1081	CTGGAATCGACACCACTCCGGAACAAGCTTGGACAGCGCGCTTAAGCTATTCAAGAG	1140
Db	12952	CTGGAATCGACACCACTCCGGAACAAGCTTGGACAGCGCGCTTAAGCTATTCAAGAG	13011
OY	1141	CAGCTGTGGGAAGGATGATATCATGTGGAGGTAAAGCGGTATTCAGAGGTCAGCGAGAGGCC	1200
Db	13012	CAGCTGTGGGAAGGATGATATCATGTGGAGGTAAAGCGGTATTCAGAGGTCAGCGAGAGGCC	13071
OY	1201	GCCCCGGCTCTGGGGGATCCAACAGAGCGCATTTGTTCCCGCTGGTGTCTACAGACGAGCTT	1260
Db	13072	GCCCCGGCTCTGGGGGATCCAACAGAGCGCATTTGTTCCCGCTGGTGTCTACAGACGAGCTT	13131
OY	1261	AACCAAGCAAGTCGTTGGTGTCACTCTGTTGCAAGGCTGGAATCTCGGTGTACACACACG	1320
Db	13132	AACCAAGCAAGTCGTTGGTGTCACTCTGTTGCAAGGCTGGAATCTCGGTGTACACACACG	13191
OY	1321	AACGAAGCTCTCGACGCTGCTGCTGGATATCATGAGCTTACAGAGACATAGGGAGCTCCGTC	1380
Db	13192	AACGAAGCTCTCGACGCTGCTGCTGGATATCATGAGCTTACAGAGACATAGGGAGCTCCGTC	13251
OY	1381	CTCGCGTGGGACATCTGTGACAGAGATGTTCCCGCGCAGACTTCTGGACGACATGCTCGAA	1440
Db	13252	CTCGCGTGGGACATCTGTGACAGAGATGTTCCCGCGCAGACTTCTGGACGACATGCTCGAA	13311
OY	1441	GCGTACGTGTTTTTCTCCGCGGCTCATCTAGAGAACCATAGGGGTGAACAGGTGCCCTGT	1500
Db	13312	GCGTACGTGTTTTTCTCCGCGGCTCATCTAGAGAACCATAGGGGTGAACAGGTGCCCTGT	13371
OY	1501	TGCGTTCCCGCTCGCCACAGCTAAGAACGCGCGGCGCAGCCGCAACCGCGCTGCTG	1560
Db	13372	TGCGTTCCCGCTCGCCACAGCTAAGAACGCGCGGCGCAGCCGCAACCGCGCTGCTG	13431
OY	1561	AGCAGACATACGCTGCACAGCGCTGTTGCGCGCGCGGCTCAGGAGCTGCCATGACAGCTC	1620
Db	13432	AGCAGACATACGCTGCACAGCGCTGTTGCGCGCGCGGCTCAGGAGCTGCCATGACAGCTC	13491
OY	1621	GCCGTGTGTCTCGGCGCGCAAGACGCTTCAAGAGAGCTTTGCGCGCGGTGTGCGGCGCA	1680
Db	13492	GCCGTGTGTCTCGGCGCGCAAGACGCTTCAAGAGAGCTTTGCGCGCGGTGTGCGGCGCA	13551
OY	1681	CTTGGCGCGCGCGCTGCCCGAGACAGGGGGGCAAGCCGCCAACCATTTGGTCGCGGTGTGATG	1740
Db	13552	CTTGGCGCGCGCGCTGCCCGAGCAGGGGGGCAAGCCGCCAACCATTTGGTCGCGGTGTGATG	13611
OY	1741	GAGAAAGGCTGGAGAGAGGTTGTGCGCGGTTCTCGCGGTGTCTGAGTCAAGCGCGGCGCTAC	1800
Db	13612	GAGAAAGGCTGGAGAGAGGTTGTGCGCGGTTCTCGCGGTGTCTGAGTCAAGCGCGGCGCTAC	13671
OY	1801	GTCGCGATTCGATGCGGACCTTACCGGCGGAGCGTATTCACCTACCTCTCGATCATGSGTGA	1860
Db	13672	GTCGCGATTCGATGCGGACCTTACCGGCGGAGCGTATTCACCTACCTCTCGATCATGSGTGA	13731
OY	1861	GTAAGAGCTCTGTCTGACAGCAGCCATGCTGTGAATGACAACTGTTCATATGCGCGCGGAGATC	1920
Db	13732	GTAAGAGCTCTGTCTGACAGCAGCCATGCTGTGAATGACAACTGTTCATATGCGCGCGGAGATC	13791
OY	1921	CAGCGGCGCTCTGTGAGAGGCGCGGCTGTGAAGAGGCAAGGAGCAACAGCTCTCGATGATG	1980
Db	13792	CAGCGGCGCTCTGTGAGAGGCGCGGCTGTGAAGAGGCAAGGAGCAACAGCTCTCGATGATG	13851
OY	1981	CCCAATTGACACACTTGCGATCTTCGCGTATGTCTACTTACACTTCGCGGATTCACAGGGTTG	2040
Db	13852	CCCAATTGACACACTTGCGGATCTTCGCGTATGTCTACTTACACTTCGCGGATTCACAGGGTTG	13911
OY	2041	CCCAAGGGGCTGATGATGATCATCGGGGTCGCTCAACCATCTCGACATCAACGAG	2100

Db	13912	CCCAAGGGGGTGAATGATCCATCAATCCGGGGTCCGCTCAACCAATCTCGAACAATCAAGAG	13974
Oy	2101	CGCTTCGAAATAGGGCCCGGAGACAGAGTGTCTGGCGCTCTCTCGCTGAGCTTTCGATCTTC	2160
Db	13972	CGCTTCGAAATAGGGCCCGGAGACAGAGTGTCTGGCGCTCTCTCGCTGAGCTTTCGATCTTC	14031
Oy	2161	TCCGCTCATATAGTGTTCGGGGATCTCGGGGGGGGGGGATACATCGTGGTGGCGAGCGCG	2220
Db	14032	TCCGCTCATATAGTGTTCGGGGATCTCGGGGGGGGGGGATACATCGTGGTGGCGAGCGCG	14091
Oy	2221	TCCAAAGCTGGCGCATCCGGCGCATTTGGGCGAGATTGATCGAAACGAGAGAAGGTGAACGTGTG	2280
Db	14092	TCCAAAGCTGGCGCATCCGGCGCATTTGGGCGAGATTGATCGAAACGAGAGAAGGTGAACGTGTG	14151
Oy	2281	TGGAACCTCGGTGCCGGCGCTGATGCGAGATCTCTCGACCATTTTGAAGGTGGCCCCGAT	2340
Db	14152	TGGAACCTCGGTGCCGGCGCTGATGCGAGATCTCTCGACCATTTTGAAGGTGGCCCCGAT	14211
Oy	2341	TCCGCTCCCTAGAGTCTCTGGCGGCTTTTGGCTCTTGAAGCGGAGCACTGGATCCCGGTGGGGCCCG	2400
Db	14212	TCCGCTCCCTAGAGTCTCTGGCGGCTTTTGGCTCTTGAAGCGGAGCACTGGATCCCGGTGGGGCCCG	14271
Oy	2401	CCTGCGCAGAGCTCCAGGCGCATCAGAGCCCGGCGTGTCCGTGATCAGCCCTTGAGCGGAGCACCC	2460
Db	14272	CTTGGCAGAGCTCCAGGCGCATCAGAGCCCGGCGTGTCCGTGATCAGCCCTTGAGCGGAGCACCC	14331
Oy	2461	GAAGCGTGCATCTGTGTCATCCGGGTACCCCGTAGGAGACGTGCACTTATCTGGGCGAGC	2520
Db	14332	GAAGCGTGCATCTGTGTCATCCGGGTACCCCGTAGGAGACGTGCACTTATCTGGGCGAGC	14391
Oy	2521	ATCCCTCAACGGCCGCTCCGCTGGCGCAACAGACCTTCCACGTCTGATAGAGAGCGCTCGAA	2580
Db	14392	ATCCCTCAACGGCCGCTCCGCTGGCGCAACAGACCTTCCACGTCTGATAGAGAGCGCTCGAA	14451
Oy	2581	CCGGGACCGGCTCTGGGTTCCGGGGGCAACTCTACATTGGCGGGGTCCGGCTGGCACTGGGC	2640
Db	14452	CCGGGACCGGCTCTGGGTTCCGGGGGCAACTCTACATTGGCGGGGTCCGGCTGGCACTGGGC	14511
Oy	2641	TACTGGCGCATGAGAGAGAGACGCGCAAGAGCTTCTGTGTGACCCCGAGACCCGGGAG	2700
Db	14512	TACTGGCGCATGAGAGAGAGACGCGCAAGAGCTTCTGTGTGACCCCGAGACCCGGGAG	14571
Oy	2701	CGCTCTCAACAAGACCGGCGATCTGGGCGCTACTCTCCCGATGGAACAATCGAGTTCAATG	2760
Db	14572	CGCTCTCAACAAGACCGGCGATCTGGGCGCTACTCTCCCGATGGAACAATCGAGTTCAATG	14631
Oy	2761	GGGCGTGAAGACCAACCAATCAAGCTTCCGCGATACCGGTTGAGCTCGGGGAAATCGAG	2820
Db	14632	GGGCGTGAAGACCAACCAATCAAGCTTCCGCGATACCGGTTGAGCTCGGGGAAATCGAG	14691
Oy	2821	GAACCGCTCAAGTCCGATCCGAACGTACCGGACGCGGTGATTTGTGCTCGGGAAACGAC	2880
Db	14692	GAACCGCTCAAGTCCGATCCGAACGTACCGGACGCGGTGATTTGTGCTCGGGAAACGAC	14751
Oy	2881	GGCGGCAACAAGCTCTTCAAGCTTATGTGTCTCCCGAAGGCAACAGAGAGAGCGGCTGGC	2940
Db	14752	GGCGGCAACAAGCTCTTCAAGCTTATGTGTCTCCCGAAGGCAACAGAGAGAGCGGCTGGC	14811
Oy	2941	GAGCAGGACCGGAGGCTCAAGACCGAGCGGATGACGCGAGAGCAACACCGCCGAAAGCG	3000
Db	14812	GAGCAGGACCGGAGGCTCAAGACCGAGCGGATGACGCGAGAGCAACACCGCCGAAAGCG	14871
Oy	3001	GACGGCTTGAAGCGACGGCGAGAGAGGTGCAATTCAAGCTCGCTCGACACGAGATCCGGAGG	3060
Db	14872	GACGGCTTGAAGCGACGGCGAGAGAGGTGCAATTCAAGCTCGCTCGACACGAGATCCGGAGG	14931
Oy	3061	GACCTGGAACGAAAGCCCGCTGCTGATCTGAACCGGACGAGATCCCGGGAGGCGGGGCTGG	3120
Db	14932	GACCTGGAACGAAAGCCCGCTGCTGATCTGAACCGGACGAGATCCCGGGAGGCGGGGCTGG	14991
Oy	3121	GACGTCTAACCGGCTCGCCCTAACGCTCCAAACCTTCTTGAAGGCCCGGATTCGGTTGTT	3180

Db 14992 GACCTGTAAGCGCGCTGCGCGTAGCGTCCGAACGTTCTTGAAGCCCGCATTCCTGTTGTT 15051
Qy 3181 GAGTTGGTTCGATTCCTGAGCTGCTTGAAGAGCGCGGACCGGACCGGACCGCTTCCC 3240
Db 15052 GAGTTGGTTCGATTCCTGAGCTGCTTGAAGAGCGCGGACCGGACCGGACCGCTTCCC 15111
Qy 3241 AAATTCGGTTATCCATTCGCGCGGAGCAAGTACCCGGTGCAGAACTTACGCGTATGCAAA 3300
Db 15112 AAATTCGGTTATCCATTCGCGCGGAGCAAGTACCCGGTGCAGAACTTACGCGTATGCAAA 15171
Qy 3301 TCCGCGCCGATTCGAGAGCGGTGAGACGAGGCGTTCTATTATTACACCCGTTGAGACCGT 3360
Db 15172 TCCGCGCCGATTCGAGAGCGGTGAGACGAGGCGTTCTATTATTACACCCGTTGAGACCGT 15231
Qy 3361 TTGCTGAAGCTCTCCGATCAGCGGATTCGAGCGCGGAGCGACGTTGCGCAAACTTGCAC 3420
Db 15232 TTGCTGAAGCTCTCCGATCAGCGGATTCGAGCGCGGAGCGACGTTGCGCAAACTTGCAC 15291
Qy 3421 GTGTTGATGAAGCGCGCTTCAACCTCTGTTCTGAGGCGAGGATCGACGCCATCGAGTGC 3480
Db 15292 GTGTTGATGAAGCGCGCTTCAACCTCTGTTCTGAGGCGAGGATCGACGCCATCGAGTGC 15351
Qy 3481 CTGATGATGATGCTGCTCGCGAGAAATTTTGCTGCTGAGAGCGCGATTAATGCGGACGTC 3540
Db 15352 CTGATGATGATGCTGCTCGCGAGAAATTTTGCTGCTGAGAGCGCGATTAATGCGGACGTC 15411
Qy 3541 CTGATGAGAGAGCGCGCTTCTCTGCAACAATCGCGCTGCTGCGGAGGCGCAATTCATTTT 3600
Db 15412 CTGATGAGAGAGCGCGCTTCTCTGCAACAATCGCGCTGCTGCGGAGGCGCAATTCATTTT 15471
Qy 3601 GAAACGTTGCGCGCGGTTCTGCACTGCGACAATTGGAAGTTTACGTGACGCGACGTCG 3660
Db 15472 GAAACGTTGCGCGCGGTTCTGCACTGCGACAATTGGAAGTTTACGTGACGCGACGTCG 15531
Qy 3661 GCGCGGCGGATGAGACCGCGGAGATTCAGAGTCTGTAGCTGTGCTGAGATTCCTCACCG 3720
Db 15532 GCGCGGCGGATGAGACCGCGGAGATTCAGAGTCTGTAGCTGTGCTGAGATTCCTCACCG 15591
Qy 3721 AGCGCGCGGAGAGCGCGGCGCGCCCTCCCGGCGCGAGACCACTTTCGCGATATGCTT 3780
Db 15592 AGCGCGCGGAGAGCGCGGCGCGCCCTCCCGGCGCGAGACCACTTTCGCGATATGCTT 15651
Qy 3781 CGCGACCTTCTTGAAGACCAAACTACCCGAGTACATGTCCTTACGCTTCTGTCGAGCTC 3840
Db 15652 CGCGACCTTCTTGAAGACCAAACTACCCGAGTACATGTCCTTACGCTTCTGTCGAGCTC 15711
Qy 3841 GATGCGTTGCGGCTGCTGCTCAACCGGCAAGTTCATGTAAGCCCTGCGGAGCGGAG 3900
Db 15712 GATGCGTTGCGGCTGCTGCTCAACCGGCAAGTTCATGTAAGCCCTGCGGAGCGGAG 15771
Qy 3901 GATACCTGTCGCGCGCGCATTCGCGGAGCAAGCGGCGCAAGCGGAGCGCTTGAAGAGATC 3960
Db 15772 GATACCTGTCGCGCGCGCATTCGCGGAGCAAGCGGCGCAAGCGGAGCGCTTGAAGAGATC 15831
Qy 3961 CTCGTCGCGGTCGTAAGGAGGTGCTCGGCTGAGAGTGTGCGGCTCCAGAGAGCTTC 4020
Db 15832 CTCGTCGCGGTCGTAAGGAGGTGCTCGGCTGAGAGTGTGCGGCTCCAGAGAGCTTC 15891
Qy 4021 GTCCATCTTGGTGCAGATTCGATTAATCGTTGCGATGAGAGCGCTTGTGCAAGAGG 4080
Db 15892 GTCCATCTTGGTGCAGATTCGATTAATCGTTGCGATGAGAGCGCTTGTGCAAGAGG 15951
Qy 4081 CTGATGAGGAGATGCGCATCAGAGTGTTCAGTACCCGAACCTGCGGCTGCGTGC 4140
Db 15952 CTGATGAGGAGATGCGCATCAGAGTGTTCAGTACCCGAACCTGCGGCTGCGTGC 16011
Qy 4141 TCCGTTTGCAGCGAGCTCGAGAGATTAATCAGCGCGCGCAACATCAGAGCCGAGTG 4200
Db 16012 TCCGTTTGCAGCGAGCTCGAGAGATTAATCAGCGCGCGCAACATCAGAGCCGAGTG 16071
Qy 4201 GAGGTTGCGCGCAAGGCGAGAGAGTGTGCTAA 4233
Db 16072 GAGGTTGCGCGCAAGGCGAGAGAGTGTGCTAA 16104

RESULT 4
AR199559
LOCUS AR199559 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6355458.
ACCESSION AR199559
VERSION AR199559.1 GI:20249633
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Liqon,J, Madison., Molnar,I., Zirkle,R., Cyr,D,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epoethlones
JOURNAL Patent: US 6355458-A 1 12-MAR-2002;
FEATURES
source
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 100.0%; Score 4233; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGACATCATTCAGCTTCTGAAAGAGCTGAGACGAGGATATCAAGCTGGCGCGAT 60
Db 11872 ATGACATCATTCAGCTTCTGAAAGAGCTGAGACGAGGATATCAAGCTGGCGCGAT 11931
Qy 61 GGGAGGCGCTCCAGATACAGGCGCCCAAGAACCCCTGAACCCGAACCTGCTGCTGA 120
Db 11932 GGGAGGCGCTCCAGATACAGGCGCCCAAGAACCCCTGAACCCGAACCTGCTGCTGA 11991
Qy 121 ATCTCCAGACAAAGACGATCTGACGATGCTCGTCGAGACATCCCGCAGAAATCC 180
Db 11992 ATCTCCAGACAAAGACGATCTGACGATGCTCGTCGAGACATCCCGCAGAAATCC 12051
Qy 181 ATCTGCGCGCGCGCGAGCGGAGCGACGCTCCGTTCTCTCAAGACATCAAGAAATCC 240
Db 12052 ATCTGCGCGCGCGCGAGCGGAGCGACGCTCCGTTCTCTCAAGACATCAAGAAATCC 12111
Qy 241 TACTGCTGGGCGGAGCAGAGCGTTTACGGTCCCAAGGAGATCAAGCTTATCGCGAA 300
Db 12112 TACTGCTGGGCGGAGCAGAGCGTTTACGGTCCCAAGGAGATCAAGCTTATCGCGAA 12171
Qy 301 TACGACTGTACGATCTCGAGCTGCGAGCGCTGAGCGCGCTTTGGAAGTGTGCGG 360
Db 12172 TACGACTGTACGATCTCGAGCTGCGAGCGCTGAGCGCGCTTTGGAAGTGTGCGG 12231
Qy 361 CGGACACGATGCTTGGGCGCCACAGCGCTGCCACATGATGAGATGATGAGCTTAA 420
Db 12232 CGGACACGATGCTTGGGCGCCACAGCGCTGCCACATGATGAGATGATGAGCTTAA 12291
Qy 421 GTACAGCGGACATCGAGATCATGATCTGCGGCGCTGACCGGAGACACAGGAAAGG 480
Db 12292 GTACAGCGGACATCGAGATCATGATCTGCGGCGCTGACCGGAGACACAGGAAAGG 12351
Qy 481 AGGCTGTGTGTTGCGAGATGCGATGTCGACCGCATTAATGACACCGAGCGCTTCG 540
Db 12352 AGGCTGTGTGTTGCGAGATGCGATGTCGACCGCATTAATGACACCGAGCGCTTCG 12411
Qy 541 CTCTATACGCTGTCGCGCTTGGCTGAGACGAGGCGAAACCGCTGCTGCTAGATC 600
Db 12412 CTCTATACGCTGTCGCGCTTGGCTGAGACGAGGCGAAACCGCTGCTGCTAGATC 12471
Qy 601 GATCTCATTAAGTTGAGCTTGAAGCGCTGCTCATCTTCAAGACTGCTGAGCTTC 660
Db 12472 GATCTCATTAAGTTGAGCTTGAAGCGCTGCTCATCTTCAAGACTGCTGAGCTTC 12531
Qy 661 TACGAAGATCCGAGACCTTCTCTGCTGTGAGGCTCTGTAACCGGATTAATGATC 720

Db 12532 TACGAAGATCCCGAAGACTCTCTCCCTGCTCGAGCTCTCGTACCCGCAATTATGACTTC 12591
QY 721 GCGCTGAGATCTCGCAAGAGTCTGAGCGCATCAACGATCGATGATTAATGGAACGG 780
Db 12592 GCGCTGAGATCTCGCAAGAGTCTGAGCGCATCAACGATCGATGATTAATGGAACGG 12651
QY 781 CCGATCGCGAGAGTCCCACTCCCGCGAGAGCTTCCGATGAGAGCCGATCCATCTACCTTC 840
Db 12652 CCGATCGCGAGAGTCCCACTCCCGCGAGAGCTTCCGATGAGAGCCGATCCATCTACCTTC 12711
QY 841 AAGAGATCCGCTTCGCGACACAGAGCATGAGCTCCGTCGGACTCTCGGGTCGATTG 900
Db 12712 AAGAGATCCGCTTCGCGACACAGAGCATGAGCTCCGTCGGACTCTCGGGTCGATTG 12771
QY 901 AAGCGCGTGTGCGGAGAGCGCGGGCTGACCCCGACGCGGCGTCACTCTGCTGCAATTTTC 960
Db 12772 AAGCGCGTGTGCGGAGAGCGCGGGCTGACCCCGACGCGGCGTCACTCTGCTGCAATTTTC 12831
QY 961 GAGTGATCGGCGCTGAGAGCGCGAGCCCGGTTTACGCTCAACATAACGCTTTCAAC 1020
Db 12832 GAGTGATCGGCGCTGAGAGCGCGAGCCCGGTTTACGCTCAACATAACGCTTTCAAC 12891
QY 1021 CCGCTCCCGTCCATCCGCGCGTGAACGATATCAACGCGGAGCTTCACTCGATGATGCTTC 1080
Db 12892 CCGCTCCCGTCCATCCGCGCGTGAACGATATCAACGCGGAGCTTCACTCGATGATGCTTC 12951
QY 1081 CTGGAATCGACACCACTGCGCAAGAGCTTGAACAGCGCGCTAAAGCTATTCAGAG 1140
Db 12952 CTGGAATCGACACCACTGCGCAAGAGCTTGAACAGCGCGCTAAAGCTATTCAGAG 13011
QY 1141 CAGCTGTGGAAGCGATGATCACTGCAAGCTAAAGGCTTACGAGTCCAGCGAGAGGCC 1200
Db 13012 CAGCTGTGGAAGCGATGATCACTGCAAGCTAAAGGCTTACGAGTCCAGCGAGAGGCC 13071
QY 1201 GCCCGGATCTCGGAGATCCACAGAGCGGATTTCCCGTGGTGTCAAGAGCGGCTT 1260
Db 13072 GCCCGGATCTCGGAGATCCACAGAGCGGATTTCCCGTGGTGTCAAGAGCGGCTT 13131
QY 1261 AACCGAAGTCTGTGCTGATCACTCTGTTGACAGGCTCGGAACTCCGCTTACACAC 1320
Db 13132 AACCGAAGTCTGTGCTGATCACTCTGTTGACAGGCTCGGAACTCCGCTTACACAC 13191
QY 1321 AGCGAATCTCTAGCTGCTGATCACTCACTGATGAGAGCATGAGGAGCCCTGCTC 1380
Db 13192 AGCGAATCTCTAGCTGCTGATCACTCACTGATGAGAGCATGAGGAGCCCTGCTC 13251
QY 1381 CTCGCTGAGGACATCTGCAAGAGTGTTCCTCCGCGACCTTCTGAGACGATGCTCGAA 1440
Db 13252 CTCGCTGAGGACATCTGCAAGAGTGTTCCTCCGCGACCTTCTGAGACGATGCTCGAA 13311
QY 1441 GCGTACGTGTTTTTCTCCGCGGCTCACTGAGAAACATGAGGATGACAGTGCCTGT 1500
Db 13312 GCGTACGTGTTTTTCTCCGCGGCTCACTGAGAAACATGAGGATGACAGTGCCTGT 13371
QY 1501 TCGCTCCGCGCTCCAGCTAGAAAGCGCGGCGAGGCAAGCAACGCGCTGCTG 1560
Db 13372 TCGCTCCGCGCTCCAGCTAGAAAGCGCGGCGAGGCAAGCAACGCGCTGCTG 13431
QY 1561 AGCGAATATCGTGAAGCGCTGTTCGCGCGCGGATCGAGAGTGCCTCACTGACCTC 1620
Db 13432 AGCGAATATCGTGAAGCGCTGTTCGCGCGCGGATCGAGAGTGCCTCACTGACCTC 13491
QY 1621 GCGGTGTGTGCGCGCAAGACGCTTACGTAAGAAAGCTTTCGCGCGTTCGCGCGCA 1680
Db 13492 GCGGTGTGTGCGCGCAAGACGCTTACGTAAGAAAGCTTTCGCGCGTTCGCGCGCA 13551
QY 1681 CTGCGCGCGGCTGCGAGAGCGGCGGCAAGCGCGCAACATGATCGCGGTGATG 1740
Db 13552 CTGCGCGCGGCTGCGAGAGCGGCGGCAAGCGCGCAACATGATCGCGGTGATG 13611
QY 1741 GAGAAAGCTGAGAGAGTGTTCGCGGTTCGCGGTGCTCGAGTCAGGCGCGGCTTAC 1800
Db 13612 GAGAAAGCTGAGAGAGTGTTCGCGGTTCGCGGTGCTCGAGTCAGGCGCGGCTTAC 13671

QY 1801 GTGCCGATCATGCCGACCTTACCGCGGAGCGGTATCCCACTACTCTCTCGATCATGTGAG 1860
Db 13672 GTGCCGATCATGCCGACCTTACCGCGGAGCGGTATCCCACTACTCTCTCGATCATGTGAG 13711
QY 1861 GTAAAGCTGTGCTGAGAGAGCATGAGTGTGAGCAACGTGCAATGAGCCGCGGGATC 1920
Db 13732 GTAAAGCTGTGCTGAGAGAGCATGAGTGTGAGCAACGTGCAATGAGCCGCGGGATC 13791
QY 1921 CAGCGGCTCTGTGAGCGAGGCGCGCTCGAAGCGACGCGGACCACTTCGATGATG 1980
Db 13792 CAGCGGCTCTGTGAGCGAGGCGCGCTCGAAGCGACGCGGACCACTTCGATGATG 13851
QY 1981 CCCATTGACACCTTCCGATCTCGCGTATGATCACTTACCTCGGATTCACAGGCTTC 2040
Db 13852 CCCATTGACACCTTCCGATCTCGCGTATGATCACTTACCTCGGATTCACAGGCTTC 13911
QY 2041 CCCAAGGGGATGATGATCATCGAGGAGCGGCTCGAAGCGACCACTTCGATGATG 2100
Db 13912 CCCAAGGGGATGATGATCATCGAGGAGCGGCTCGAAGCGACCACTTCGATGATG 13971
QY 2101 CGCTTCAAAATAGGCGCGGAGACAGGAGTCTGCGCTCTCTCGCTGAGCTTCATCTC 2160
Db 13972 CGCTTCAAAATAGGCGCGGAGACAGGAGTCTGCGCTCTCTCGCTGAGCTTCATCTC 14031
QY 2161 TCGCTCTATGATGTGTTGGGATCTCTGCGCGCGGCGTAAAGATGATGCGGACGCG 2220
Db 14032 TCGCTCTATGATGTGTTGGGATCTCTGCGCGCGGCGTAAAGATGATGCGGACGCG 14091
QY 2221 TCCAGCTGCGGATCCGCGGATGAGGATGAGGATGATGCAAGAGAGATGACGCTG 2280
Db 14092 TCCAGCTGCGGATCCGCGGATGAGGATGAGGATGATGCAAGAGAGATGACGCTG 14151
QY 2281 TCGAACTCGGTCCGCGCTGATGCGATGCTCTGTGAGCAATTTGAGGCTGCCGAT 2340
Db 14152 TCGAACTCGGTCCGCGCTGATGCGATGCTCTGTGAGCAATTTGAGGCTGCCGAT 14211
QY 2341 TCGCTCGCTAGTCTCTGCGGCTTTCGCTGAGCGCGGCTGATCCGCTGGGCGCTG 2400
Db 14212 TCGCTCGCTAGTCTCTGCGGCTTTCGCTGAGCGCGGCTGATCCGCTGGGCGCTG 14271
QY 2401 CCGGCGAGCTTCAGGCAATCAGGCGCGGCGGCTGCTGATGATGAGCTGAGCGGCGCAC 2460
Db 14272 CCGGCGAGCTTCAGGCAATCAGGCGCGGCGGCTGCTGATGATGAGCTGAGCGGCGCAC 14331
QY 2461 GAAAGCTGATGTGCTCATGCGGATACCCCTGAGAAACGTGCACTTTCGTGCGGAGC 2520
Db 14332 GAAAGCTGATGTGCTCATGCGGATACCCCTGAGAAACGTGCACTTTCGTGCGGAGC 14391
QY 2521 ATCCCTACGCGCGCTCCGCTGCGCAACGAGCTTCCAGCTGATGAGGCGCTCGAA 2580
Db 14392 ATCCCTACGCGCGCTCCGCTGCGCAACGAGCTTCCAGCTGATGAGGCGCTCGAA 14451
QY 2581 CCGCGCGGCTGTGAGTTCCGCGGCACTTACATTTGCGGAGTCCGAGCTGCACTGAGC 2640
Db 14452 CCGCGCGGCTGTGAGTTCCGCGGCACTTACATTTGCGGAGTCCGAGCTGCACTGAGC 14511
QY 2641 TACTGAGCGATGAGAGAGAGCGCGCAAGACTTCTGTGTCACCCGAGACCGGGAG 2700
Db 14512 TACTGAGCGATGAGAGAGAGCGCGCAAGACTTCTGTGTCACCCGAGACCGGGAG 14571
QY 2701 CCGCTCTACAGACCGGCGGATGTTGGCGGCTTACCTGCGCGAGGAAACATGAGTTCA 2760
Db 14572 CCGCTCTACAGACCGGCGGATGTTGGCGGCTTACCTGCGCGAGGAAACATGAGTTCA 14631
QY 2761 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATACCGGCTTGAAGCTGCGGAAATCGAG 2820
Db 14632 GGGCGTGAAGCAACCAATCAAGCTTCCGCGATACCGGCTTGAAGCTGCGGAAATCGAG 14691
QY 2821 GAAAGCTTCAAGTCCGATTCGAACTGCGGACGCGGATGATGCTGCGGAAACGAG 2880
Db 14692 GAAAGCTTCAAGTCCGATTCGAACTGCGGACGCGGATGATGCTGCGGAAACGAG 14751

QY	2881	GGCGCGAAACAAGCTCTTCTAGCGCTATGTGTGCCCGAGGGACAAGAGAAGCGCTGCC	2940
Db	14752	GGCGCGAAACAAGCTCTTCTAGCGCTATGTGTGCCCGAGGGACAAGAGAAGCGCTGCC	14811
QY	2941	GAGCAGGACGCGAGCCTCAAGAACCGAGCGGATTCGACGCGAGACAACGCGCGGAAAGCG	3000
Db	14812	GAGCAGGACGCGAGCCTCAAGAACCGAGCGGATTCGACGCGAGACAACGCGCGGAAAGCG	14871
QY	3001	GACGCGCTTGAAGCCAGCGGCGAGAGGGTGCAGTTCAAGCTGCTTGACAACGCACTTCGAGAG	3060
Db	14872	GACGCGCTTGAAGCCAGCGGCGAGAGGGTGCAGTTCAAGCTGCTTGACAACGCACTTCGAGAG	14931
QY	3061	GACCTGGAACGAAAACCCGTCGTCGATCTCAACGCGGACGAGATCCGCGGGAGAGCGGGGCTG	3120
Db	14932	GACCTGGAACGAAAACCCGTCGTCGATCTCAACGCGGACGAGATCCGCGGGAGAGCGGGGCTG	14991
QY	3121	GACGCTCAACGCGCGTCGCGCTAGACGATCCGAACGTTCTTGAAGGCCCGAATTCGGTTGTT	3180
Db	14992	GACGCTCAACGCGCGTCGCGCTAGACGATCCGAACGTTCTTGAAGGCCCGAATTCGGTTGTT	15051
QY	3181	GAGTTTGGTGGATTCTTGACCTGCTTGAAGCGGTGAGCCGACGCGCGACCTTCCC	3240
Db	15052	GAGTTTGGTGGATTCTTGACCTGCTTGAAGCGGTGAGCCGACGCGCGACCTTCCC	15111
QY	3241	AAATTCGGTATCCATCGGCGGGGACAGACGTACCCGGTCGAAACCTACCGGTATGCAAA	3300
Db	15112	AAATTCGGTATCCATCGGCGGGGACAGACGTACCCGGTCGAAACCTACCGGTATGCAAA	15171
QY	3301	TCCGCGCGCATCGAGGGCGTGCAGAGAGGCTTCTATTATTCACCCGTTGACGACCGT	3360
Db	15172	TCCGCGCGCATCGAGGGCGTGCAGAGAGGCTTCTATTATTCACCCGTTGACGACCGT	15231
QY	3361	TTGCTGAGACTCTCCGATCACGCGGATCAGCGGAGCCGACGTTGCGCAAACTTCGAC	3420
Db	15232	TTGCTGAGACTCTCCGATCACGCGGATCAGCGGAGCCGACGTTGCGCAAACTTCGAC	15291
QY	3421	GTGTTGCATGAACGGGGGCTCAACTCTCCGTGTGTGGGACAGATTCGACCCATCCAGTGC	3480
Db	15292	GTGTTGCATGAACGGGGGCTCAACTCTCCGTGTGTGGGACAGATTCGACCCATCCAGTGC	15351
QY	3481	CTGATATGATCGTTCGTCGCGAGAAATTTGCTGCTGAGAGCGCGATATATGCGCGACGTC	3540
Db	15352	CTGATATGATCGTTCGTCGCGAGAAATTTGCTGCTGAGAGCGCGATATATGCGCGACGTC	15411
QY	3541	CTGATGAGAGAGCGGCGCTTCCTGCAACATCGGCGTCTGTCCGGTGGGCAATTCATATTT	3600
Db	15412	CTGATGAGAGAGCGGCGCTTCCTGCAACATCGGCGTCTGTCCGGTGGGCAATTCATATTT	15471
QY	3601	GAACAGGTTGGGCGCGGTTCTTCGACCTGCGACATTTGCGACGTTTACGTGCAACGCAATGCTG	3660
Db	15472	GAACAGGTTGGGCGCGGTTCTTCGACCTGCGACATTTGCGACGTTTACGTGCAACGCAATGCTG	15531
QY	3661	GGCGGGGGGGTGAAGCCCGGGGCAAGTTCCAGTCTGTACGTCGGGTGAGGATTCCTCACCG	3720
Db	15532	GGCGGGGGGGTGAAGCCCGGGGCAAGTTCCAGTCTGTACGTCGGGTGAGGATTCCTCACCG	15591
QY	3721	AGGCGCGCACGACGCGCGCGCGCCCTTCGCGCGCGAGACGACTTCGCGCATATGCTT	3780
Db	15592	AGGCGCGCACGACGCGCGCGCGCCCTTCGCGCGCGAGACGACTTCGCGCATATGCTT	15651
QY	3781	CGCGACTTCTTGAAGACCAACTACCCGAGTACATGGTGCCTTACAGTCTTTCGAGGCTC	3840
Db	15652	CGCGACTTCTTGAAGACCAACTACCCGAGTACATGGTGCCTTACAGTCTTTCGAGGCTC	15711
QY	3841	GATGCGGTTGGCCGCTGAAGTCCAAACGGAAGGTGCAGTGTAAAGGCCCTTGCGCAGCGGAAG	3900
Db	15712	GATGCGGTTGGCCGCTGAAGTCCAAACGGAAGGTGCAGTGTAAAGGCCCTTGCGCAGCGGAAG	15771
QY	3901	GATACCTTCGTGCGCGCGGCAATTCGAGGACACAGCGCCCAACGCGACCTTCGAGAGATC	3960
Db	15772	GATACCTTCGTGCGCGCGGCAATTCGAGGACACAGCGCCCAACGCGACCTTCGAGAGATC	15831
QY	3961	CTGCTCGCGGTCGTACGGAAGGTCTCGGCTGGAAGTGTGCGGCTCCAGCAGAGCTTC	4020

Db	15832	CTGTCGGGTTCGACGGAGGTGTCTCGGGCTCGAGTGTGTGGGCTTCAGCAGAGCTTC	15891
Qy	4021	GTCGATCTTGTGTGCGACATCGATTCACATCGTTCGATAGGAGGCTGTTCAGAGAGAG	4080
Db	15892	GTCGATCTTGTGTGCGACATCGATTCACATCGTTCGATAGGAGGCTGTTCAGAGAGAG	15955
Qy	4081	CTGGATAGGAGATTCGCATCACCAGAGTTGTTCCAGTACCCCGAACTTGGCTGCGTGGCG	4140
Db	15952	CTGGATAGGAGATTCGCATCACCAGAGTTGTTCCAGTACCCCGAACTTGGCTGCGTGGCG	16011
Qy	4141	TCCGGTTTGGCCCGAGACTTCGAGAGATCTAGATCAACGGCCGAGCAATGACAGAACCGAGTG	4200
Db	16012	TCCGGTTTGGCCCGAGACTTCGAGAGATCTAGATCAACGGCCGAGCAATGACAGAACCGAGTG	16071
Qy	4201	GAGGTTGGGCGCAAGGCGAGAGACGTAGCTAA	4233
Db	16072	GAGGTTGGGCGCAAGGCGAGAGACGTAGCTAA	16104

RESULT	5
LOCUS	AR199567
DEFINITION	Sequence 1 from patent US 6355459.
ACCESSION	AR199567
VERSION	AR199567.1 GI:20249641
KEYWORDS	.
SOURCE	. Unknown.
ORGANISM	. Unknown.
REFERENCE	Unclassified. 1 (bases 1 to 68750) <i>Schupp,T., Higon,J.Madison., Molnar,I., Zitzke,R., Cyr,D.Dawn. and Gorlach,J.</i> Genes for the biosynthesis of epoethilones Patent: US 6355459-A 1 12-MAR-2002; Location/Qualifiers location 1..68750 /organism="unknown"
TITLE	JOURNAL
FEATURES	source
BASE COUNT	9596 a 22456 c 25359 g 11159 t
ORIGIN	

Query Match	Best Local Similarity	100.0%;	Score 4233;	DB 6;	Length 68750;
Matches 4233;	Conservative	0;	Mismatches	0;	Indels
		0;	Gaps	0;	
QY	1	ATGACGATCAATACGCTTTCTGAAICGAGCTCGAGCACCCAGGGATATCAAGTCGGCGCCGAT	60		
Db	11872	ATGACGATCAATACGCTTTCTGAAICGAGCTCGAGCACCCAGGGATATCAAGTCGGCGCCGAT	11931		
QY	61	GGGGAGGCGCTCCAGATPACAGGCGCCCAAGAAAGCGCTTAACCCGAACTGGCTCGCTGAA	120		
Db	11932	GGGGAGGCGCTCCAGATPACAGGCGCCCAAGAAAGCGCTTAACCCGAACTGGCTCGCTGAA	11991		
QY	121	ATCTCCAGACAAAAAGCAGATCTTGAGATGCTTCCTCAGAGATCTCCCGCAGATTC	180		
Db	11992	ATCTCCAGACAAAAAGCAGATCTTGAGATGCTTCCTCAGAGATCTCCCGCAGATTC	12051		
QY	181	ATCGTCCCGCCCGCAGCGCGGACGCTCCGTTTCTCTCAGACATCCAAAGATTC	240		
Db	12052	ATCGTCCCGCCCGCAGCGCGGACGCTCCGTTTCTCTCAGACATCCAAAGATTC	12111		
QY	241	TACTGCGTGGGCGGACAGGAGGGTTTACGATCCCGAGGGGATCCAGACCTATCGCGAA	300		
Db	12112	TACTGCGTGGGCGGACAGGAGGGTTTACGATCCCGAGGGGATCCAGACCTATCGCGAA	12171		
QY	301	TACGACTGTACGAGATCTCGACGTCGCGAGGCTGAGCGCGGCTTTGGAAGTCGTGCG	360		
Db	12172	TACGACTGTACGAGATCTCGACGTCGCGAGGCTGAGCGCGGCTTTGGAAGTCGTGCG	12231		
QY	361	CGGACGACATGCTTGGGGCCCAACGCTGCCGACATATGACAGTGTATCAGGCTTAA	420		
Db	12232	CGGACGACATGCTTGGGGCCCAACGCTGCCGACATATGACAGTGTATCAGGCTTAA	12291		

OY	421	GTGACGCCGACATCGAGATCATCGATCTGCGGGCTGACCGGAGCACAGGGAAAGG	480
Db	12292	GTGACGCCGACATCGAGATCATCGATCTGCGGGCTGACCGGAGCACAGGGAAAGG	12351
OY	481	AGGCTCGTGTGCTTGCGAGATGCGATGTGCGCACCGCATCTATGACACGAGCCCTCGG	540
Db	12352	AGGCTCGTGTGCTTGCGAGATGCGATGTGCGCACCGCATCTATGACACGAGCCCTCGG	12411
OY	541	CTCTATCAGCTGCTGCGCGTTCCGGCTGGACGAGCGGCAACCCTGTCTGTGCTCAGTATC	600
Db	12412	CTCTATCAGCTGCTGCGCGTTCCGGCTGGACGAGCGGCAACCCTGTCTGTGCTCAGTATC	12471
OY	601	GATTCCTATTAACGTTGACCTTAGCAGCGCTTCCATCATCTTCAAGGACTGCGTCACTTC	660
Db	12472	GATTCCTATTAACGTTGACCTTAGCAGCGCTTCCATCATCTTCAAGGACTGCGTCACTTC	12531
OY	661	TACGAAGATCCCGAGACCTCTCTCCCTGCTGTGAGCTCTGACCGCGATTATGATATC	720
Db	12532	TACGAAGATCCCGAGACCTCTCTCCCTGCTGTGAGCTCTGACCGCGATTATGATATC	12591
OY	721	GCCTGTGAGTCTCGCAAGAACTGAGCGGCATCAACGATCGATGATTACTGGAAGCGG	780
Db	12592	GCCTGTGAGTCTCGCAAGAACTGAGCGGCATCAACGATCGATGATTACTGGAAGCGG	12651
OY	781	CGCATTCGCCAGCTCTCCCACTTCGCGCGACGCTTCGATGAAGCGCATCATCTACCTCG	840
Db	12652	CGCATTCGCCAGCTCTCCCACTTCGCGCGACGCTTCGATGAAGCGCATCATCTACCTCG	12711
OY	841	AAGAGATATCCGCTTCGCGCACAGAGCATTGAGCTGCGCGCATCTCGGGGTGCAATTG	900
Db	12712	AAGAGATATCCGCTTCGCGCACAGAGCATTGAGCTGCGCGCATCTCGGGGTGCAATTG	12771
OY	901	AAGCGGCGTGTCCGGGAGCGCGGCTGACCCCGACGCGCTCATCTCGCTGATTTTTC	960
Db	12772	AAGCGGCGTGTCCGGGAGCGCGGCTGACCCCGACGCGCTCATCTCGCTGATTTTTC	12831
OY	961	GAGGTGATTCGGGGCTGTGAGCGCGAGGCCCGCTTACGTCAACATACGCTTCAAC	1020
Db	12832	GAGGTGATTCGGGGCTGTGAGCGCGAGGCCCGCTTACGTCAACATACGCTTCAAC	12891
OY	1021	CGGCTCCCGGTCATCCGGCGGTGAAGGATATCACGGGGACTTCAGCTGATGCTC	1080
Db	12892	CGGCTCCCGGTCATCCGGCGGTGAAGGATATCACGGGGACTTCAGCTGATGCTC	12951
OY	1081	CTGGACATTCGACACCACTCGCGACAAAGGCTTCGAAAGGCGGCTTAAGGATTTCAAGG	1140
Db	12952	CTGGACATTCGACACCACTCGCGACAAAGGCTTCGAAAGGCGGCTTAAGGATTTCAAGG	13011
OY	1141	CAGCTGTGGGAGCGATGATCATCTGCGAGTAAAGCGGTATCGAGGTCCAGCGAGGCTC	1200
Db	13012	CAGCTGTGGGAGCGATGATCATCTGCGAGTAAAGCGGTATCGAGGTCCAGCGAGGCTC	13071
OY	1201	GCCCGGGGCTCGGGGGATTCACAAGAGGGGCAATTGTTCCCGGTGTCTCAAGCGGCGCTT	1260
Db	13072	GCCCGGGGCTCGGGGGATTCACAAGAGGGGCAATTGTTCCCGGTGTCTCAAGCGGCGCTT	13131
OY	1261	AACCAAGCAAGTCTGTGATGTCACCTCTGTTCAGAGGCTCGGAATCCCGGTATACACAGC	1320
Db	13132	AACCAAGCAAGTCTGTGATGTCACCTCTGTTCAGAGGCTCGGAATCCCGGTATACACAGC	13191
OY	1321	ACGCAAGACTCTCAAGCTGTCTGTGATCATCACTCACTTACAGACGATGGGACCTCGTC	1380
Db	13192	ACGCAAGACTCTCAAGCTGTCTGTGATCATCACTCACTTACAGACGATGGGACCTCGTC	13251
OY	1381	CTGCGGCGGAGCATCGTCGACGAGAGTGTCCGCGCGGACCTTCTGGAACGATGCTCGAA	1440
Db	13252	CTGCGGCGGAGCATCGTCGACGAGAGTGTCCGCGCGGACCTTCTGGAACGATGCTCGAA	13311
OY	1441	GCCTAGCTGTGTTTTCTCCGGCGGCTCACTGAGAAACCATAGGGTAAACAGGTGCGCTGT	1500
Db	13312	GCCTAGCTGTGTTTTCTCCGGCGGCTCACTGAGAAACCATAGGGTAAACAGGTGCGCTGT	13371
OY	1501	TGCGTTCGCGCTGCGCAAGCTAGAAAGCGCGGCGAGCGCAACGCGACCAACGCGCTGCTG	1560

Db	13372	TCGCTTTCCGCTTGCCCAAGCTAAGAGCGCGGAGCGCAACCGCAGCTC	13433
Qy	1561	AGCGAGCATACGCTGCACGGGCTTTTCGCGCGCGGGGTGAGCAGCTGCCATGCAGCTC	1620
Db	13432	AGCGAGCATACGCTGCACGGGCTTTTCGCGCGCGGGGTGAGCAGCTGCCATGCAGCTC	13491
Qy	1621	GCCGTGATTCGCGCGCGCAAGAGCTCAGTACGAAGAGCTTTCCGCGCGGTTCGCGCGCA	1680
Db	13492	GCCGTGATTCGCGCGCGCAAGAGCTCAGTACGAAGAGCTTTTCGCGCGGTTCGCGCGCA	13551
Qy	1681	CTTGCGCGCGGGCGCGCGAGCAGGGGGGACCGCCGCAACATTTGGTCGCGGNGATG	1740
Db	13552	CTTGCGCGCGGGCGCGCGAGCAGGGGGGACCGCCGCAACATTTGGTCGCGGNGATG	13611
Qy	1741	GAGAAAGCTGGGAGCAGTTGTTCGCGGTTCTCGCGGTCTCAGTCAGCGCGGCTAC	1800
Db	13612	GAGAAAGCTGGGAGCAGTTGTTCGCGGTTCTCGCGGTCTCAGTCAGCGCGGCTAC	13671
Qy	1801	GTCGCATTCGATGCGGACCTTACCGCGCGAGCGTATTCACCTCTCGATCATGTGTAG	1860
Db	13672	GTCGCATTCGATGCGGACCTTACCGCGCGAGCGTATTCACCTCTCGATCATGTGTAG	13731
Qy	1861	GTAAGCTTCGTGTCAGCGGCAATGGGTGGAGGGAACCTGTCATAGGCGCGCGGGATC	1920
Db	13732	GTAAGCTTCGTGTCAGCGGCAATGGGTGGAGGGAACCTGTCATAGGCGCGCGGGATC	13791
Qy	1921	CAGCGGCTGCTCGTGAAGCAGGCGCGGCTCGAAGCGAGCGGACCGCTCGATGATG	1980
Db	13792	CAGCGGCTGCTCGTGAAGCAGGCGCGGCTCGAAGCGAGCGGACCGCTCGATGATG	13851
Qy	1981	CCCATTCAGACACCTTCGATCTTCGCGTATGATTCACCTCGGGATTCACAGGGTTG	2040
Db	13852	CCCATTCAGACACCTTCGATCTTCGCGTATGATTCACCTCGGGATTCACAGGGTTG	13911
Qy	2041	CCCAAGGGGGTGATGATTCGATCATCGGGGGCGGTCAACACATCTGGAACATCAAGAG	2100
Db	13912	CCCAAGGGGGTGATGATTCGATCATCGGGGGCGGTCAACACATCTGGAACATCAAGAG	13971
Qy	2101	CGCTTCGAAATAGGGGCGCGGAGACAGGGGTCTGGCGCTCTCGCTGAGCTTTCGATCTC	2160
Db	13972	CGCTTCGAAATAGGGGCGCGGAGACAGGGGTCTGGCGCTCTCGCTGAGCTTTCGATCTC	14031
Qy	2161	TCGGTCTATGATGTTCGGGAATCTTGCGCGCGGGCGGTACGATGTCGTGCCGACGCG	2220
Db	14032	TCGGTCTATGATGTTCGGGAATCTTGCGCGCGGGCGGTACGATGTCGTGCCGACGCG	14091
Qy	2221	TCCAACTGCGGCAATCCGCGGCAATTTGGGACAGATTTATTCGAACGAAAGAGGACGGT	2280
Db	14092	TCCAACTGCGGCAATCCGCGGCAATTTGGGACAGATTTATTCGAACGAAAGAGGACGGT	14151
Qy	2281	TGGAATCTCGGTGCGCGGCTGATAGCGGAATCTGTCGAGACATTTTAGGGTCGCCCGAT	2340
Db	14152	TGGAATCTCGGTGCGCGGCTGATAGCGGAATCTGTCGAGACATTTTAGGGTCGCCCGAT	14211
Qy	2341	TCGCTGCTAGGTCTCTCGCGGCTTTTCGCTGCTGAGCGGCGACTGATCCCGGTGGGCTTG	2400
Db	14212	TCGCTGCTAGGTCTCTCGCGGCTTTTCGCTGCTGAGCGGCGACTGATCCCGGTGGGCTTG	14271
Qy	2401	CTGCGGAGGATCCAGGCGCAACAGGCCCGGCGGTGCGGTATACGCTGGGGCGGGGACAC	2460
Db	14272	CTGCGGAGGATCCAGGCGCAACAGGCCCGGCGGTGCGGTATACGCTGGGGCGGGGACAC	14331
Qy	2461	GAAGCTGTCATTCGTGTCATCGGGTACCCGCTGAGAGACGTGCACTTATCTGTGGGAGAC	2520
Db	14332	GAAGCTGTCATTCGTGTCATCGGGTACCCGCTGAGAGACGTGCACTTATCTGTGGGAGAC	14391
Qy	2521	ATCCCTTACGCGCGGTTCGCTGCTGCGCAACAGACGTTTACGCTGTCATGAGCGCTCGAA	2580
Db	14392	ATCCCTTACGCGCGGTTCGCTGCTGCGCAACAGACGTTTACGCTGTCATGAGCGCTCGAA	14451
Qy	2581	CCGCGCGCGGTTCGTGGTTCCGGGGCACTTACATTTGGCGGGTCCGGCTGCGCATGGGC	2640

Db 14452 CCGGCCCCGCTTGGGTTCCGGGCACTCTAATTTGGGGGCTGGGCTGACCTGGGC 14511
Qy 2641 TACTGGCGCGATGAAGAGAGACGCGCAAGAGCTTCTCTGTCACCCCGAGACCGGGAG 2700
Db 14512 TACTGGCGCGATGAAGAGAGACGCGCAAGAGCTTCTCTGTCACCCCGAGACCGGGAG 14571
Qy 2701 CGCCTTACAGACCGGCGATCTGGGCGCTACCTGCCCCGATGAAACATCGAGTTCAAG 2760
Db 14572 CGCCTTACAGACCGGCGATCTGGGCGCTACCTGCCCCGATGAAACATCGAGTTCAAG 14631
Qy 2761 GGGCGTGAAGCAACCAATCAAGTTGGCGGATACCCCGTTGAGCTGGGGAAATCGAG 2820
Db 14632 GGGCGTGAAGCAACCAATCAAGTTGGCGGATACCCCGTTGAGCTGGGGAAATCGAG 14691
Qy 2821 GAAACGCTCAAGTGCATCCGACGTACGCGACGCGGTGATTTGTCGCCCTGGAGAACAC 2880
Db 14692 GAAACGCTCAAGTGCATCCGACGTACGCGACGCGGTGATTTGTCGCCCTGGAGAACAC 14751
Qy 2881 GCGGCGAACCAAGCTCTTCTAGCCCTATGTTGTCGGAAGGACACGAGACGCGCTGC 2940
Db 14752 GCGGCGAACCAAGCTCTTCTAGCCCTATGTTGTCGGAAGGACACGAGACGCGCTGC 14811
Qy 2941 GAGCAGACGCGAGCTTCAAAACCGAGGAGATCCAGCGGAGACACACCGCCGAGACG 3000
Db 14812 GAGCAGACGCGAGCTTCAAAACCGAGGAGATCCAGCGGAGACACACCGCCGAGACG 14871
Qy 3001 GACGCGCTTGAAGCGAGCGGAGGAGGATTCAGAGCTCGCTCGACACGAGCTCCGAGG 3060
Db 14872 GACGCGCTTGAAGCGAGCGGAGGAGGATTCAGAGCTCGCTCGACACGAGCTCCGAGG 14931
Qy 3061 GACCTGACGGAAGACCCGTCGTGATCTGAACCGGACAGATCCGCGGAGAGCGGGGCTG 3120
Db 14932 GACCTGACGGAAGACCCGTCGTGATCTGAACCGGACAGATCCGCGGAGAGCGGGGCTG 14991
Qy 3121 GACGCTTACGCGCGCTTCCGAGAGCTCCGAACTCTTGAAGCCCGATTCCTGTTGTT 3180
Db 14992 GACGCTTACGCGCGCTTCCGAGAGCTCCGAACTCTTGAAGCCCGATTCCTGTTGTT 15051
Qy 3181 GAGTTTGGTGCATTCCTGAGGCTTGTGACAGCGTGAGCCCGAGCGGCGACCCCTTCC 3240
Db 15052 GAGTTTGGTGCATTCCTGAGGCTTGTGACAGCGTGAGCCCGAGCGGCGACCCCTTCC 15111
Qy 3241 AAATTCGCTTATCCATCGCGGCGGACGACGTAACCCGCTGCAAACTTACGCTATGTCAA 3300
Db 15112 AAATTCGCTTATCCATCGCGGCGGACGACGTAACCCGCTGCAAACTTACGCTATGTCAA 15171
Qy 3301 TCCGCGCCGATCGAGGCGGTGAGACGAGGCTTCTATTATACACCGCTTGAAGACCT 3360
Db 15172 TCCGCGCCGATCGAGGCGGTGAGACGAGGCTTCTATTATACACCGCTTGAAGACCT 15231
Qy 3361 TTGCTGAAGCTCTCCGATCAAGGAGATGAGGCGGAGGCACTTCCGCAAACTTGCAC 3420
Db 15232 TTGCTGAAGCTCTCCGATCAAGGAGATGAGGCGGAGGCACTTCCGCAAACTTGCAC 15291
Qy 3421 GTGTTTCATGAAGCGGCGCTTCAACTCTGTTTCGTGCGGACAGATGACGCCATCGAGTC 3480
Db 15292 GTGTTTCATGAAGCGGCGCTTCAACTCTGTTTCGTGCGGACAGATGACGCCATCGAGTC 15351
Qy 3481 CTGATGAGATCGTGTCCGAGAAATTTTCTGCTCTGAGAGCCGATATATGCGCAGCTC 3540
Db 15352 CTGATGAGATCGTGTCCGAGAAATTTTCTGCTCTGAGAGCCGATATATGCGCAGCTC 15411
Qy 3541 CTGATGAGAGCGGCGCTTCCGCAACATCGGCGCTGTCGCGTGGGGCAATTCATTTT 3600
Db 15412 CTGATGAGAGCGGCGCTTCCGCAACATCGGCGCTGTCGCGTGGGGCAATTCATTTT 15471
Qy 3601 GAACAGGTTCCGCGCGTTCTGACCTTCGACATTCGACGTTTACGTGACGCGCATGCTG 3660
Db 15472 GAACAGGTTCCGCGCGTTCTGACCTTCGACATTCGACGTTTACGTGACGCGCATGCTG 15531
Qy 3661 GCGCGGCGGATGACCGCGGCGAGTTCCAGGTTCTTACGCTCGGTGAGAGATTCCTCACCG 3720
Db 15532 GCGCGGCGGATGACCGCGGCGAGTTCCAGGTTCTTACGCTCGGTGAGAGATTCCTCACCG 15591

Qy 3721 AGGCGCGCAGACGCGCGGCGCCCTCCCGCCGAGACGACCTTCCGATATGCTT 3780
Db 15592 AGGCGCGCAGACGCGCGGCGCCCTCCCGCCGAGACGACCTTCCGATATGCTT 15651
Qy 3781 CGCAGCTTCTTGAAGACCAACTACCGGATGATGCTGCTTACAGCTTCTGAGGCTC 3840
Db 15652 CGCAGCTTCTTGAAGACCAACTACCGGATGATGCTGCTTACAGCTTCTGAGGCTC 15711
Qy 3841 GATGCTTCCCGCTGAGCTCCCAAGGAGTGCATGATGAGGCGCTGCGGAGCGAG 3900
Db 15712 GATGCTTCCCGCTGAGCTCCCAAGGAGTGCATGATGAGGCGCTGCGGAGCGAG 15771
Qy 3901 GATACCTGCTGCGCGCGCACTTGGGCGACACGCGCACGCGGACGCTTGGAGAGATC 3960
Db 15772 GATACCTGCTGCGCGCGCACTTGGGCGACACGCGCACGCGGACGCTTGGAGAGATC 15831
Qy 3961 CTGCTCGGAGTCTGATCGGAGAGTCTCGGCTGAGAGTGTGCGGCTTCAGACGCTTC 4020
Db 15832 CTGCTCGGAGTCTGATCGGAGAGTCTCGGCTGAGAGTGTGCGGCTTCAGACGCTTC 15891
Qy 4021 GTGATCTTGTGAGATGATGATTCATGATGCTTGCATGAGAGGCTTTCAGAGAG 4080
Db 15892 GTGATCTTGTGAGATGATGATTCATGATGCTTGCATGAGAGGCTTTCAGAGAG 15951
Qy 4081 CTGATAGGAGATGCGCATACCGAGTTGTTCCAGTACCGAACTTCGCTCGCTGCGG 4140
Db 15952 CTGATAGGAGATGCGCATACCGAGTTGTTCCAGTACCGAACTTCGCTCGCTGCGG 16011
Qy 4141 TCCGCTTTCGCGGAGACTGAGAGATCTAGATACGCGCGGAACTGACGAGCCGAGTG 4200
Db 16012 TCCGCTTTCGCGGAGACTGAGAGATCTAGATACGCGCGGAACTGACGAGCCGAGTG 16071
Qy 4201 GAGGTTGCGCGAGGCGGAGAGAGCTAGCTAA 4233
Db 16072 GAGGTTGCGCGAGGCGGAGAGAGCTAGCTAA 16104

RESULT 6
AR201097
LOCUS AR201097 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6358719-A 1 19-MAR-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 100.0%; Score 4233; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGACATCATCATGCTTCTGAAAGAGCTGAGACCAAGGATATCAAGCTGGCGCGAT 60
Db 11872 ATGACATCATCATGCTTCTGAAAGAGCTGAGACCAAGGATATCAAGCTGGCGCGAT 11931
Qy 61 GGGAGAGCGCTCCAGATACAGGCGCCCAAGAACGCGCTGAAACCGGAACCTGCTGCTGA 120
Db 11932 GGGAGAGCGCTCCAGATACAGGCGCCCAAGAACGCGCTGAAACCGGAACCTGCTGCTGA 11991
Qy 121 ATCTCGAGACCAAAAGACGATCTGACGATGCTTCGTCAGAGACTTCCCGCAGAAATCC 180

Db 11992 ATCTCCGAGCACAAAAGCAGCATCTGACGATGCTCCGTACAGACTCCCGGAGAAATCC 12051
Qy 181 ATCTGTCCTCCCGCCAGCCGAGCGGACGCTCCGTTTCTCTCTCAACAGATCTCAAGATTC 240
Db 12052 ATCTGTCCTCCCGCCAGCCGAGCGGACGCTCCGTTTCTCTCTCAACAGATCTCAAGATTC 12111
Qy 241 TACTGCTGGGCGGACAGAGGCTTTCAGTCTCCCGAGGAGGATCCAGCCTATCGGAA 300
Db 12112 TACTGCTGGGCGGACAGAGGCTTTCAGTCTCCCGAGGAGGATCCAGCCTATCGGAA 12171
Qy 301 TACGACTGTACGGAATCTGACGTCGCGAGGCTGAGCCGCTTTTGAAAGTCGTGCG 360
Db 12172 TACGACTGTACGGAATCTGACGTCGCGAGGCTGAGCCGCTTTTGAAAGTCGTGCG 12231
Qy 361 CCGGACAGATATCTTGGGCGCCACAGCTGCGGAGATATGATGAGGTATTCAGACTTAA 420
Db 12232 CCGGACAGATATCTTGGGCGCCACAGCTGCGGAGATATGATGAGGTATTCAGACTTAA 12291
Qy 421 GTCCAGCGCCGACATCGAGATCATCGATCTGCGCGGCTGCAACCGGAGCAACGGGAGGCG 480
Db 12292 GTCCAGCGCCGACATCGAGATCATCGATCTGCGCGGCTGCAACCGGAGCAACGGGAGGCG 12351
Qy 481 AGGCTGTGTGCTTGGGAGATGCGATGTGCAACCGCATCTATGACACCGAGCGCCCTCCG 540
Db 12352 AGGCTGTGTGCTTGGGAGATGCGATGTGCAACCGCATCTATGACACCGAGCGCCCTCCG 12411
Qy 541 CTCTATCAGCTGTGCTGCGCTTGGGCTGACAGCGGCAAAACCGCTCTGCTGCTCAATTC 600
Db 12412 CTCTATCAGCTGTGCTGCGCTTGGGCTGACAGCGGCAAAACCGCTCTGCTGCTCAATTC 12471
Qy 601 GATCTCATTTAAGCTTGAACCTTGAAGCAGCTGTCCATCATCTTGAAGGCTGAGCTCACTTC 660
Db 12472 GATCTCATTTAAGCTTGAACCTTGAAGCAGCTGTCCATCATCTTGAAGGCTGAGCTCACTTC 12531
Qy 661 TACGAGAGATCCGAGACCTCTCTCCCTGTCTGAGAGCTCTCGTACCGGATTAATGATCTC 720
Db 12532 TACGAGAGATCCGAGACCTCTCTCCCTGTCTGAGAGCTCTCGTACCGGATTAATGATCTC 12591
Qy 721 GGGCTGGAATCTTCGCAAGAGTCTGAGGCTGATCAACGATTCGATTTACTGGAAGCGG 780
Db 12592 GGGCTGGAATCTTCGCAAGAGTCTGAGGCTGATCAACGATTCGATTTACTGGAAGCGG 12651
Qy 781 CCGATGCGCGAGCTCCACCTCCGCGGACGCTTCGATGAAGGCGGATCATCTACCTTCG 840
Db 12652 CCGATGCGCGAGCTCCACCTCCGCGGACGCTTCGATGAAGGCGGATCATCTACCTTCG 12711
Qy 841 AAGAGATCCGCTTCCGGAACAAGAGCAATGCTGCGATCTCTGAGGCTGATTC 900
Db 12712 AAGAGATCCGCTTCCGGAACAAGAGCAATGCTGCGATCTCTGAGGCTGATTC 12771
Qy 901 AAGCGCGGTGTCGGGAGCGCGGCTGACCCGAGCGGCTGATCTCTGAGCTGATTTTC 960
Db 12772 AAGCGCGGTGTCGGGAGCGCGGCTGACCCGAGCGGCTGATCTCTGAGCTGATTTTC 12831
Qy 961 GAGGTGATCGGCGCTGAGAGCGGAGCCCGGCTTACCTCAACATTAACGCTTTCAAC 1020
Db 12832 GAGGTGATCGGCGCTGAGAGCGGAGCCCGGCTTACCTCAACATTAACGCTTTCAAC 12891
Qy 1021 CCGGCTCCCGCTTCATCCGCGCTGAAAGATATCAACGGGAGCTTCAAGTGTCTTC 1080
Db 12892 CCGGCTCCCGCTTCATCCGCGCTGAAAGATATCAACGGGAGCTTCAAGTGTCTTC 12951
Qy 1081 CTGGAATGGAACCACTCGGCAACAAGACTTGAACAGCGGCTTAAGGTATTCAGAG 1140
Db 12952 CTGGAATGGAACCACTCGGCAACAAGACTTGAACAGCGGCTTAAGGTATTCAGAG 13011
Qy 1141 CAGCTTGGGAAGCGATGATCACTGCGAGATTAAGGCTATCGAGGTCAGGAGGCGC 1200
Db 13012 CAGCTTGGGAAGCGATGATCACTGCGAGATTAAGGCTATCGAGGTCAGGAGGCGC 13071
Qy 1201 GCCCGGGTCTTGGAATCCACGAGGCGCATTTGCTCCGCTGCTGCTACGAGCGGCTT 1260
Db 13072 GCCCGGGTCTTGGAATCCACGAGGCGCATTTGCTCCGCTGCTGCTACGAGCGGCTT 13131

Qy 1261 AACGAGAGTCTGTTGATCACTCTGTTGAGAGGCTCGGAACCTCCGCTGACACGAC 1320
Db 13132 AACGAGAGTCTGTTGATCACTCTGTTGAGAGGCTCGGAACCTCCGCTGACACGAC 13191
Qy 1321 ACGGAGCTCTCAGCTCTGCTGATCATGACTCTACGAGCAGATGAGGAGCTCTGTC 1380
Db 13192 ACGGAGCTCTCAGCTCTGCTGATCATGACTCTACGAGCAGATGAGGAGCTCTGTC 13251
Qy 1381 CTGCGGTGGAATGTGTGAGCGGAGTGTCCCGCCGACCTTTGAGAGCATGTCTCGAA 1440
Db 13252 CTGCGGTGGAATGTGTGAGCGGAGTGTCCCGCCGACCTTTGAGAGCATGTCTCGAA 13311
Qy 1441 GCGTACGTGTTTTCTCCGCGGCTCACTGAGAACCATGGGAGTGAACAGGTGCTGT 1500
Db 13312 GCGTACGTGTTTTCTCCGCGGCTCACTGAGAACCATGGGAGTGAACAGGTGCTGT 13371
Qy 1501 TCGCTTCCGCTCCAGCTAGAAAGCGGCGGAGCGGAAACGCGACCAACGCGCTGCTG 1560
Db 13372 TCGCTTCCGCTCCAGCTAGAAAGCGGCGGAGCGGAAACGCGACCAACGCGCTGCTG 13431
Qy 1561 ACGGAGCATACGCTGACAGGCTGTTCGCGCGCGGCTGAGCAGCTGCCATGCACTTC 1620
Db 13432 ACGGAGCATACGCTGACAGGCTGTTCGCGCGCGGCTGAGCAGCTGCCATGCACTTC 13491
Qy 1621 GCGGAGTGTGCGCGCGGAGAGCGCTCACTGAGAAAGCTTTGCGCGCTTCCGCGGGA 1680
Db 13492 GCGGAGTGTGCGCGCGGAGAGCGCTCACTGAGAAAGCTTTGCGCGCTTCCGCGGGA 13551
Qy 1681 CTGCGCGCGGCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1740
Db 13552 CTGCGCGCGGCTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 13611
Qy 1741 GAGAAAGCTGAGAGCAAGTGTGCGCGTCTCCGCTGCTGAGTCAAGGCGCGGCTTAC 1800
Db 13612 GAGAAAGCTGAGAGCAAGTGTGCGCGTCTCCGCTGCTGAGTCAAGGCGCGGCTTAC 13671
Qy 1801 GTCGCGATGAGAGCGGAGCTTACCGGCGGAGGAGTATCACTACCTCTGATCATGAGTGA 1860
Db 13672 GTCGCGATGAGAGCGGAGCTTACCGGCGGAGGAGTATCACTACCTCTGATCATGAGTGA 13731
Qy 1861 GTAAAGCTGTGCTGACGAGCAGCAGTGTGATGAGCAACTGTCAATGAGCGCGGAGATC 1920
Db 13732 GTAAAGCTGTGCTGACGAGCAGCAGTGTGATGAGCAACTGTCAATGAGCGCGGAGATC 13791
Qy 1921 CAGCGGCTGCTGTGAGAGGAGCGGCGCTTCAAGGCGAGCGGAGCAGCTTCCATGATG 1980
Db 13792 CAGCGGCTGCTGTGAGAGGAGCGGCGCTTCAAGGCGAGCGGAGCAGCTTCCATGATG 13851
Qy 1981 CCGATTCAAGACCTTCGATGCTGCTGATGATCATCTACACTCGGAGATCCACAGGTTG 2040
Db 13852 CCGATTCAAGACCTTCGATGCTGCTGATGATCATCTACACTCGGAGATCCACAGGTTG 13911
Qy 2041 CCGAAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100
Db 13912 CCGAAGGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 13971
Qy 2101 CCGTTGAAATGAGGCGCGGAGACAGGCTGCTGCTCTCTCCCTGAGCTTTCATCTTC 2160
Db 13972 CCGTTGAAATGAGGCGCGGAGACAGGCTGCTGCTCTCTCCCTGAGCTTTCATCTTC 14031
Qy 2161 TCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2220
Db 14032 TCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 14091
Qy 2221 TCGAAGCTGCGGATCCGCGGATGAGGAGATGATGATGATGATGATGATGATGATGATG 2280
Db 14092 TCGAAGCTGCGGATCCGCGGATGAGGAGATGATGATGATGATGATGATGATGATGATG 14151
Qy 2281 TCGAAGCTGCGGATCCGCGGATGAGGAGATGATGATGATGATGATGATGATGATGATG 2340
Db 14152 TCGAAGCTGCGGATCCGCGGATGAGGAGATGATGATGATGATGATGATGATGATGATG 14211

QY	2341	TCGCTCCGCTAGAGTCTCTCGAGGCTTTGGCTCTGAGCGGCGACCTGAAATCCCGGAGGCGCTG	2400
Db	14212	TCGCTCCGCTAGAGTCTCTCGAGGCTTTGGCTCTGAGCGGCGACCTGAAATCCCGGAGGCGCTG	14271
QY	2401	CTTGCGAGGCTCCAGGCGCATCAGGCGCGGCGGTGTCGGTGTATCAGCCTGGGCGGGCCACC	2460
Db	14272	CTTGCGAGGCTCCAGGCGCATCAGGCGCGGCGGTGTCGGTGTATCAGCCTGGGCGGGCGCACCC	14333
QY	2461	GAAGCGTTCGATCTGGTCCATTCGGGTACCCGTGAGGAAGTGCACCTATCTGTGGCGAGC	2520
Db	14332	GAAGCGTTCGATCTGGTCCATTCGGGTACCCGTGAGGAAGTGCACCTATCTGTGGCGAGC	14391
QY	2521	ATCCGCTACGAGCGGTCCGCTGGCGCAACAGACGTTCCAGGTGCTCATGAGGCGGCTCGAA	2580
Db	14392	ATCCGCTACGAGCGGTCCGCTGGCGCAACAGACGTTCCAGGTGCTCATGAGGCGGCTCGAA	14451
QY	2581	CCGCGCCCGGCTCTGGGTCTCGGGGCAACTGATTCGCGGGGTCTGGGCTGCGACTGGGCG	2640
Db	14452	CCGCGCCCGGCTCTGGGTCTCGGGGCAACTGATTCGCGGGGTCTGGGCTGCGACTGGGCG	14511
QY	2641	TACTGGCGCGATGAGAGGAAGACGCGCAGAGCTTCTCTGTGCACCCCGAGACCGGGGAG	2700
Db	14512	TACTGGCGCGATGAGAGGAAGACGCGCAGAGCTTCTCTGTGCACCCCGAGACCGGGGAG	14571
QY	2701	CGGCTTCACAAAGCCGGGCGATCTGGGGCGGTACCTCGCCGAGATGGAACATTCAGATTCAAG	2760
Db	14572	CGGCTTCACAAAGCCGGGCGATCTGGGGCGGTACCTCGCCGAGATGGAACATTCAGATTCAAG	14631
QY	2761	GGGCGGTAGAGCAACAACAAATCAAGCTTCGGGAGTACCGCGTTGAGCTCGGGGAAATCGAG	2820
Db	14632	GGGCGGTAGAGCAACAACAAATCAAGCTTCGGGAGTACCGCGTTGAGCTCGGGGAAATCGAG	14691
QY	2821	GAAACGCTCAAGTGCATCCGAAAGTACGCGACCGGTGATTGTGCCCTCGTGGGAAAGCAG	2880
Db	14692	GAAACGCTCAAGTGCATCCGAAAGTACGCGACCGGTGATTGTGTGCCCTCGTGGGAAAGCAG	14751
QY	2881	GCGCGCAACAAGCTCTTCTTAGGCTATGTGTGTCCCGAGGGGCAACAGGAGACGCGTGTGCC	2940
Db	14752	GCGCGCAACAAGCTCTTCTTAGGCTATGTGTGTCCCGAGGGGCAACAGGAGACGCGTGTGCC	14811
QY	2941	GAGCAGAGACCGAGCCTCCAGAGACCGAGCGGATCGACGCGAGACACACGCGCGCGAAGGCG	3000
Db	14812	GAGCAGAGACCGAGCCTCCAGAGACCGAGCGGATCGACGCGAGACACACGCGCGCGAAGGCG	14871
QY	3001	GACGGCTTGAACGACGCGGAGAGGGGTGCAAGTTCGCTCGACACGAGACTTCCGAGAG	3060
Db	14872	GACGGCTTGAACGACGCGGAGAGGGGTGCAAGTTCGCTCGACACGAGACTTCCGAGAG	14931
QY	3061	GACCTGGAACGAAAGCCCGTCTGTGATCTGACCGGCGAGGATCCGCGAGGCGGGGCTG	3120
Db	14932	GACCTGGAACGAAAGCCCGTCTGTGATCTGACCGGCGAGGATCCGCGAGGCGGGGCTG	14991
QY	3121	GACGTCACGCGGTGCGCGTAGGTCGGAACGTTCCGTTGAGGCGCCCGAATTCGGTTGTT	3180
Db	14992	GACGTCACGCGGTGCGCGTAGGTCGGAACGTTCCGTTGAGGCGCCCGAATTCGGTTGTT	15051
QY	3181	GAGTTTGGTGCATCTCGAGCTGCTTGAAGCAGCGGTGAGCCGACAGCGGCGACCTTCCC	3240
Db	15052	GAGTTTGGTGCATCTCGAGCTGCTTGAAGCAGCGGTGAGCCGACAGCGGCGACCTTCCC	15111
QY	3241	AAATTCCGTTATTCATTCGCGGAGCAGACGTAACCCGGTGCMAACCTACGCGTATGTCAAA	3300
Db	15112	AAATTCCGTTATTCATTCGCGGAGCAGACGTAACCCGGTGCMAACCTACGCGTATGTCAAA	15171
QY	3301	TCCGCGCGGATTCAGAGGGCGGTGACAGAGGCTTCTATTATTAACAACCGTTTCAGGACCGCT	3360
Db	15172	TCCGCGCGGATTCAGAGGGCGGTGACAGAGGCTTCTATTATTAACAACCGTTTCAGGACCGCT	15231
QY	3361	TTGCTGAAGCTTTCGATCAACGGGATCGACGCGGAGCGCAGCTTTCGGCAAAACTTCGAC	3420
Db	15232	TTGCTGAAGCTTTCGATCAACGGGATCGACGCGGAGCGCAGCTTTCGGCAAAACTTCGAC	15291
QY	3421	GTTGTCATGAAGCGGCGTTCAACCTCTGTTCTGTGGGACGAGATCGACCGCATCGAGTGC	3480

Db	15292	GTGTTGCATGATAAGCGGGTTCAACCTTCCTGTTGTTGGGCAAGATCGACGGCATCGAGTGG	1535		
Qy	3481	CTGTATGGAATCGTGTGCGAGAAATTTGCTGTCTGAGGCGCGATATATGCGCAGCTC	3540		
Db	15352	CTGTATGATATGTCGTGTGCGAGAAATTTTGCTGTCTGAGGCGCGAATATAGCGCAGCTC	1541		
Qy	3541	CTGTATGAGACAGGCGCTTCTCTGCAACATGTGGGTCTGTCTCGGTGGGGCAATTCATTTT	3600		
Db	15412	CTGTATGAGACAGGCGCTTCTCTGCAACATGTGGGTCTGTCTCGGTGGGGCAATTCATTTT	1547		
Qy	3601	GAACAGGTTGCGGCGGTTCTCGACTGTGCAATTCGGAAGTTTACGTGACAGGATGCTGTG	3660		
Db	15472	GAACAGGTTGCGGCGGTTCTCGACTGTGCAATTCGGAAGTTTACGTGACAGGATGCTGTG	1553		
Qy	3661	GGCGGGCGGGTGAAGACCCGCGGCAAGTTCCAGGTCTGTACGCTCGGTACAGATTCTCACC	3720		
Db	15532	GGCGGGCGGGTGAAGACCCGCGGCAAGTTCCAGGTCTGTACGCTCGGTACAGATTCTCACC	1559		
Qy	3721	AGGCGCGCCACGACGCGCGCGGCCCCCTCCGGCCGAGCAGCACTTCCGCGATATGCTT	3780		
Db	15592	AGGCGCGCCACGACGCGCGCGGCCCCCTCCGGCCGAGCAGCACTTCCGCGATATGCTT	1565		
Qy	3781	CGGCACTTCTTGAAGACCAAACTACCCGAGTACATGGTGTGCTTACAGTCTTCGTAGAGCTC	3840		
Db	15652	CGGCACTTCTTGAAGACCAAACTACCCGAGTACATGGTGTGCTTACAGTCTTCGTAGAGCTC	1571		
Qy	3841	GATGCGTTGCGCGTGAAGCTCAACGCGCAAGTGCATGTGTAAAGCCCTGCGCGAGCGGAG	3900		
Db	15712	GATGCGTTGCGCGTGAAGCTCAACGCGCAAGTGCATGTGTAAAGCCCTGCGCGAGCGGAG	1577		
Qy	3901	GATACCTGTGTCGCGCGCGGCAATTTGGGGCACACGCGCCACCGGACCGCTTGGAGAGATC	3960		
Db	15772	GATACCTGTGTCGCGCGCGGCAATTTGGGGCACACGCGCCACCGGACCGCTTGGAGAGATC	1583		
Qy	3961	CTGTCGCGGGTCCGTACGCGGAGGTGTCTCGGGCTGGAAGTGGTCCGGGCTCCAGCAGACTTC	4020		
Db	15832	CTGTCGCGGGTCCGTACGCGGAGGTGTCTCGGGCTGGAAGTGGTCCGGGCTCCAGCAGACTTC	1589		
Qy	4021	GTCGATCTTGTGTGCGACATGCATGCATTCACATGCTTCGACTGAGAGCCCTGTTGCAGAGAG	4080		
Db	15892	GTCGATCTTGTGTGCGACATGCATGCATTCACATGCTTCGACTGAGAGCCCTGTTGCAGAGAG	1595		
Qy	4081	CTGATATGAGGAGATCGGCATCACCGAGTTGTTCCAGTACCCGAACTTCGGCTCGTGGCG	4140		
Db	15952	CTGATATGAGGAGATCGGCATCACCGAGTTGTTCCAGTACCCGAACTTCGGCTCGTGGCG	1601		
Qy	4141	TCCGATTGGGCGCGAGACTCGAAGAGTCTAGATCAGCGGCGCGAACATGCAAGACCGAGTG	4200		
Db	16012	TCCGATTGGGCGCGAGACTCGAAGAGTCTAGATCAGCGGCGCGAACATGCAAGACCGAGTG	1607		
Qy	4201	GAGGTTGCGGCGCAAGGCGCAGAGACGTAGCTAA 4233			
Db	16072	GAGGTTGCGGCGCAAGGCGCAGAGACGTAGCTAA 16104			
RESULT 7					
AR208671		68750 bp	DNA	linear	PAT 20-JUN-2002
LOCUS	AR208671				
DEFINITION	Sequence 1 from patent US 6383787.				
ACCESSION	AR208671				
VERSION	AR208671.1				
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 68750)				
AUTHORS	Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.				
TITLE	Genes for the biosynthesis of epoethlones				
JOURNAL	Patent: US 6383787-A 1 07-MAY-2002;				
FEATURES	Location/Qualifiers				
SOURCE	1..68750				

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN /organism="unknown"
Query Match 100.0%; Score 4233; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGAGATCAATACGCTTCTGAACGAGCTCGACACAGGATATCAAGCTGGCGCGAT 60
11872 ATGAGATCAATACGCTTCTGAACGAGCTCGACACAGGATATCAAGCTGGCGCGAT 11931
61 GGGGAGCGCTCCAGATACAGGCGCCCAAGAAAGCCCTGAACCCGACCTGCTCGTCA 120
11932 GGGGAGCGCTCCAGATACAGGCGCCCAAGAAAGCCCTGAACCCGACCTGCTCGTCA 11991
121 ATCTCGAGACAAAAGACAGATCTCGAGATGCTCCGTCAAGAGATCCCGCAGAGATC 180
11992 ATCTCGAGACAAAAGACAGATCTCGAGATGCTCCGTCAAGAGATCCCGCAGAGATC 12051
181 ATCTGCGCGCGCGCGAGCGGACAGCTCGCTTCTCTCAAGACATCCAGAGATCC 240
12052 ATCTGCGCGCGCGCGAGCGGACAGCTCGCTTCTCTCAAGACATCCAGAGATCC 12111
241 TACTGCTGGGCGGACAGAGCGTTTACGGTCCCGAGGGATCCAGCGCTATCGGAA 300
12112 TACTGCTGGGCGGACAGAGCGTTTACGGTCCCGAGGGATCCAGCGCTATCGGAA 12171
301 TACGACTGACGAGATCTCGACGTGCGAGGCTGAGCGCGCTTTCGAAAGTCTGCGG 360
12172 TACGACTGACGAGATCTCGACGTGCGAGGCTGAGCGCGCTTTCGAAAGTCTGCGG 12231
361 CGGACAGACATGCTGGGCGGACAGCGTCCCGACATATGAGGATCGAGCTTAA 420
12232 CGGACAGACATGCTGGGCGGACAGCGTCCCGACATATGAGGATCGAGCTTAA 12291
421 GTGAGCGCGGACATGAGATCAATCGATCGCGGGCTGACCGGAGCACAGGGAAAGG 480
12292 GTGAGCGCGGACATGAGATCAATCGATCGCGGGCTGACCGGAGCACAGGGAAAGG 12351
481 AGGCTGCTGCTTGGGAGATGAGATGTCACACCGCATATGACACCGAGCGCTCCG 540
12352 AGGCTGCTGCTTGGGAGATGAGATGTCACACCGCATATGACACCGAGCGCTCCG 12411
541 CTCTATCAGCTGCTGCGCTTGGGCTGAGACAGCGGCAAAACCGCTTCTGCTCAGTATC 600
12412 CTCTATCAGCTGCTGCGCTTGGGCTGAGACAGCGGCAAAACCGCTTCTGCTCAGTATC 12471
601 GATTCATTAAGTTGAGCTAGGACGCTGTCATATCTTCAAGGACTGGGCTCAGCTTC 660
12472 GATTCATTAAGTTGAGCTAGGACGCTGTCATATCTTCAAGGACTGGGCTCAGCTTC 12531
661 TACGAGATCCGAGACCTTCTCCCTGCTCGAGACTCTGTAACCGCATTAATGATCTC 720
12532 TACGAGATCCGAGACCTTCTCCCTGCTCGAGACTCTGTAACCGCATTAATGATCTC 12591
721 GCGCTGAGATCTCGCAAGAGTCTGAGCGGACATCAAGATCGATGATTAATGAGCGG 780
12592 GCGCTGAGATCTCGCAAGAGTCTGAGCGGACATCAAGATCGATGATTAATGAGCGG 12651
781 GCGATGCGGAGCTCCACCTCGCGGAGGCTTCGATGAAGGCGGATCATCTACCTG 840
12652 GCGATGCGGAGCTCCACCTCGCGGAGGCTTCGATGAAGGCGGATCATCTACCTG 12711
841 AAGAGATCCGCTTCCGAGACAGAGCATGCTGCGCTCGACTCTGAGGCTGATG 900
12712 AAGAGATCCGCTTCCGAGACAGAGCATGCTGCGCTCGACTCTGAGGCTGATG 12771
901 AAGCGGCTGCTGGGAGCGCGGCTGACCCCGAGCGGCTCATCTGCTGATTTTC 960
12772 AAGCGGCTGCTGGGAGCGCGGCTGACCCCGAGCGGCTCATCTGCTGATTTTC 12831
961 GAGGTATGGGCGCTGAGCGGAGCGCGGCTTACGCTCAACATTAAGCTTTCAAC 1020

12832 GAGGTATGGGCGCTGAGCGGAGCGCGGCTTACGCTCAACATTAAGCTTTCAAC 12891
1021 CGGCTCCCGCTTCATCCGCGGCTGAGAGATATCACCGGAGACTTCAAGTGTCTTC 1080
12892 CGGCTCCCGCTTCATCCGCGGCTGAGAGATATCACCGGAGACTTCAAGTGTCTTC 12951
1081 CTGGAATCGACACCACTCGGAGCAAGAGCTTTCGAAACAGCGGCTTAAGCTTCAAG 1140
12952 CTGGAATCGACACCACTCGGAGCAAGAGCTTTCGAAACAGCGGCTTAAGCTTCAAG 13011
1141 CAGCTGGGAGAGCATGATATCACTGCAAGCTTAAGCGGTATGAGGTCCAGAGAGCC 1200
13012 CAGCTGGGAGAGCATGATATCACTGCAAGCTTAAGCGGTATGAGGTCCAGAGAGCC 13071
1201 GCCCGGATCTGGGAGATCAACGAGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTT 1260
13072 GCCCGGATCTGGGAGATCAACGAGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTT 13131
1261 AACGAGAAATGCTGTGTGTACCTCTGTTGCAAGAGCTCGGAACTCGGCTGTAACAGC 1320
13132 AACGAGAAATGCTGTGTGTACCTCTGTTGCAAGAGCTCGGAACTCGGCTGTAACAGC 13191
1321 ACGGAGCTCTCAGCTGCTGATGATCATGAGCTCTACGAGACAGATGGGAGCTCTGTC 1380
13192 ACGGAGCTCTCAGCTGCTGATGATCATGAGCTCTACGAGACAGATGGGAGCTCTGTC 13251
1381 CTGCGTGGAGATGTCGACGAGAGTGTCCCGCGGACCTTCTGAGACAGATGCTGAA 1440
13252 CTGCGTGGAGATGTCGACGAGAGTGTCCCGCGGACCTTCTGAGACAGATGCTGAA 13311
1441 GGTACGCTGCTTTTCTCCGCGGCTCACTAGAGAACATGGGGGAAACAGTGGCTGT 1500
13312 GGTACGCTGCTTTTCTCCGCGGCTCACTAGAGAACATGGGGGAAACAGTGGCTGT 13371
1501 TCGCTTCGCGCTCCAGCTAGAAAGCGGCGGAGCGGAAACGAGCAACGCGCTGCTG 1560
13372 TCGCTTCGCGCTCCAGCTAGAAAGCGGCGGAGCGGAAACGAGCAACGCGCTGCTG 13431
1561 AGCGAGCATGCTGTCAGCGGCTGTTGCGGCGGAGTTCGAGACGCTCCCATGCACTC 1620
13432 AGCGAGCATGCTGTCAGCGGCTGTTGCGGCGGAGTTCGAGACGCTCCCATGCACTC 13491
1621 GCGCTGCTGCTGGGCGGAGAGCGCTCAAGAGAGCTTTCGCGCGCTTTCGCGGAG 1680
13492 GCGCTGCTGCTGGGCGGAGAGCGCTCAAGAGAGCTTTCGCGCGCTTTCGCGGAG 13551
1681 CTGCGCGGCTCCGAGAGCGGAGCGGCGGAGCGGCAACATGCTGCGGCTGATG 1740
13552 CTGCGCGGCTCCGAGAGCGGAGCGGCGGAGCGGCAACATGCTGCGGCTGATG 13611
1741 GAGAAAGCTGGAGACAGGTTGTCCGGTTCTCCGCTGCTGAGTCAAGCGGCTTAC 1800
13612 GAGAAAGCTGGAGACAGGTTGTCCGGTTCTCCGCTGCTGAGTCAAGCGGCTTAC 13671
1801 GTGCCGATCGATGCGGACCTACCGGCGGAGGATACCTACCTCTGATATGATGAG 1860
13672 GTGCCGATCGATGCGGACCTACCGGCGGAGGATACCTACCTCTGATATGATGAG 13731
1861 GTAAAGCTGCTGACAGCAGCATGCTGATGCAACTGTCACTGAGCGCGGAGATC 1920
13732 GTAAAGCTGCTGACAGCAGCATGCTGATGCAACTGTCACTGAGCGCGGAGATC 13791
1921 CAGCGGCTGCTGAGCGAGCGGCGCTCGAAGGCGAGCGGACAGCTTCGATGATG 1980
13792 CAGCGGCTGCTGAGCGAGCGGCGCTCGAAGGCGAGCGGACAGCTTCGATGATG 13851
1981 CCCATTGAGACCTTTCGATGCTGCGGTATGATCACTCACTCGGAGATCCAGAGGTTG 2040
13852 CCCATTGAGACCTTTCGATGCTGCGGTATGATCACTCACTCGGAGATCCAGAGGTTG 13911
2041 CCCAAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100

Db 13912 CCCAAGGGGGTGAATATGCATATCGGGGTGCCCTCAACACCATCTCGACATCAACGAG 13971
Qy 2101 GCCTTGAATAATAGGCGCCGAGACAGAGGTGTGCGGCTCTCCTCGCTGACTTCGATCTC 2160
Db 13972 GCCTTGAATAATAGGCGCCGAGACAGAGGTGTGCGGCTCTCCTCGCTGACTTCGATCTC 14031
Qy 2161 TCGGTCTATGATGTGTGGGATCTCTGCGGCGGGCGGTACGATCTGTGTGCGGACGCG 2220
Db 14032 TCGGTCTATGATGTGTGGGATCTCTGCGGCGGGCGGTACGATCTGTGTGCGGACGCG 14091
Qy 2221 TCCAGGCTCGCGGATCCCGCGCATTTGGGAGAGTTGATTCGACGAGAGAGGTGACGGTG 2280
Db 14092 TCCAGGCTCGCGGATCCCGCGCATTTGGGAGAGTTGATTCGACGAGAGAGGTGACGGTG 14151
Qy 2281 TGAACCTGCGTCCGCGCTGATGCGGATGCTCGTGCAGCATTTTGAAGGGTGCCTCCGAT 2340
Db 14152 TGAACCTGCGTCCGCGCTGATGCGGATGCTCGTGCAGCATTTTGAAGGGTGCCTCCGAT 14211
Qy 2341 TCGCTCGCTAGGTCTCTGCGGCTTTCTGCTGAGCGCGGACCTGATCCCGGTGGGCTTG 2400
Db 14212 TCGCTCGCTAGGTCTCTGCGGCTTTCTGCTGAGCGCGGACCTGATCCCGGTGGGCTTG 14271
Qy 2401 CCTGCGAGCTCCAGGCGCATCAAGCCCGGCGCTGTGCTGATATAGCCTGCGGCGGCGCAC 2460
Db 14272 CCTGCGAGCTCCAGGCGCATCAAGCCCGGCGCTGTGCTGATATAGCCTGCGGCGGCGCAC 14331
Qy 2461 GAAGGCTGATCTGGTTCATCCGGGTACCCCGTGAAGAAAGTGCACCTATCGTGGGCGAGC 2520
Db 14332 GAAGGCTGATCTGGTTCATCCGGGTACCCCGTGAAGAAAGTGCACCTATCGTGGGCGAGC 14391
Qy 2521 ATCCCTTAAGGCGCGTCCGCTGCGCAACAGAGTTCAAGTCTCGATGAGCGCTCGAA 2580
Db 14392 ATCCCTTAAGGCGCGTCCGCTGCGCAACAGAGTTCAAGTCTCGATGAGCGCTCGAA 14451
Qy 2581 CCGGCGCGGCTGTGGGTTCCGGGCGCACTTACATTTGGCGGGGTGGGCTGCGACTGGGC 2640
Db 14452 CCGGCGCGGCTGTGGGTTCCGGGCGCACTTACATTTGGCGGGGTGGGCTGCGACTGGGC 14511
Qy 2641 TACTGGCGGATGAG 2700
Db 14512 TACTGGCGGATGAG 14571
Qy 2701 CGCCTTCAAGAGCGCGGATCTGGGCGGCTACCTGCGCGATGAGAAACATCGAGTTCAAG 2760
Db 14572 CGCCTTCAAGAGCGCGGATCTGGGCGGCTACCTGCGCGATGAGAAACATCGAGTTCAAG 14631
Qy 2761 GGGCGGTGAGAGACCAAAATCAAGCTTGGCGGATACCGGCTTGGAGTGGGAAATCGAG 2820
Db 14632 GGGCGGTGAGAGACCAAAATCAAGCTTGGCGGATACCGGCTTGGAGTGGGAAATCGAG 14691
Qy 2821 GAAACGCTCAAGTGCATCCGAAAGTACGAGCGGCGGTGATGTCGCCGTGGGAAAGAG 2880
Db 14692 GAAACGCTCAAGTGCATCCGAAAGTACGAGCGGCGGTGATGTCGCCGTGGGAAAGAG 14751
Qy 2881 GCGGCGAAAGAGCTCTTCTAGCTATGTGTCCCGAGGAGCAACAGAGAGCGGCTGCC 2940
Db 14752 GCGGCGAAAGAGCTCTTCTAGCTATGTGTCCCGAGGAGCAACAGAGAGCGGCTGCC 14811
Qy 2941 GAGCAGAGAGCGAGCTTCAAGACGAGCGGATTCGACGCGAGAGCAACGCGCGCAAGCG 3000
Db 14812 GAGCAGAGAGCGAGCTTCAAGACGAGCGGATTCGACGCGAGAGCAACGCGCGCAAGCG 14871
Qy 3001 GACGCGTTGAGCGAGCGGAGAGGGTGCAGTTCAGAGCTCGCTCGACAGGATCCGGAGG 3060
Db 14872 GACGCGTTGAGCGAGCGGAGAGGGTGCAGTTCAGAGCTCGCTCGACAGGATCCGGAGG 14931
Qy 3061 GACCTGAGCGAGAAAGCGGTCTGATGATGACCGGCGAGATCCGCGGAGCGGCGGCTG 3120
Db 14932 GACCTGAGCGAGAAAGCGGTCTGATGATGACCGGCGAGATCCGCGGAGCGGCGGCTG 14991
Qy 3121 GACGCTTACGCGCGCTGCGGCTGAGCTTCGAAAGCTTCTTGAAGCGCCGATTCGTTGTT 3180
Db 14992 GACGCTTACGCGCGCTGCGGCTGAGCTTCGAAAGCTTCTTGAAGCGCCGATTCGTTGTT 15051

Qy 3181 GAGTTGTGATTCCTGAGCTGTGAGCAGCGTGAAGCCGACGCGCGACCCCTTCC 3240
Db 15052 GAGTTGTGATTCCTGAGCTGTGAGCAGCGTGAAGCCGACGCGCGACCCCTTCC 15111
Qy 3241 AAATTCGTTATCCATCGCGCGGAGCAGCTACCGCGGTGCAAACTTACGCTATGTCAA 3300
Db 15112 AAATTCGTTATCCATCGCGCGGAGCAGCTACCGCGGTGCAAACTTACGCTATGTCAA 15171
Qy 3301 TCCGCGCGCATCGAGGGGTGGAAGAGGGCTTATATTAATCAACCGCTTGAGACCGT 3360
Db 15172 TCCGCGCGCATCGAGGGGTGGAAGAGGGCTTATATTAATCAACCGCTTGAGACCGT 15231
Qy 3361 TTGCTGAAGCTCTCCGATCAAGGATCGAGCGGAGCGACGTTCCGCAAACTTCGAC 3420
Db 15232 TTGCTGAAGCTCTCCGATCAAGGATCGAGCGGAGCGACGTTCCGCAAACTTCGAC 15291
Qy 3421 GTGTTGATGAAGCGGCTTCAACCTCTGTTCTGGGAGAGATGACGCCATCGAGTGG 3480
Db 15292 GTGTTGATGAAGCGGCTTCAACCTCTGTTCTGGGAGAGATGACGCCATCGAGTGG 15351
Qy 3481 CTGATGATGATGCTGCGGAGAAATTTGCTGCTGAGGCGGAGATATAGCGGAGCTC 3540
Db 15352 CTGATGATGATGCTGCGGAGAAATTTGCTGCTGAGGCGGAGATATAGCGGAGCTC 15411
Qy 3541 CTGATGAGCAGGCGCTTCTGCAACATCGGCGTCTGCTGCGTGGGCGAAATTCATTTT 3600
Db 15412 CTGATGAGCAGGCGCTTCTGCAACATCGGCGTCTGCTGCGTGGGCGAAATTCATTTT 15471
Qy 3601 GAACAGTTTCGCGGCTTCTGACCTGCGACATTCGAGCTTTACGTTACGTGACGCGATGCTG 3660
Db 15472 GAACAGTTTCGCGGCTTCTGACCTGCGACATTCGAGCTTTACGTTACGTGACGCGATGCTG 15531
Qy 3661 GCGCGCGGCTGAGACCCCGCGGAGTTCCAGGCTCTGACGCTCGGCAAGATTCCTCAACG 3720
Db 15532 GCGCGCGGCTGAGACCCCGCGGAGTTCCAGGCTCTGACGCTCGGCAAGATTCCTCAACG 15591
Qy 3721 AGCGCGCGCAGACGCGCGGCGGCTTCCGCGCGGAGCAGCACTTGGCGATATGCTT 3780
Db 15592 AGCGCGCGCAGACGCGCGGCGGCTTCCGCGCGGAGCAGCACTTGGCGATATGCTT 15651
Qy 3781 CGCGACTTCTTGAAGACCAAACTACCGGATGATGCTGCTCAAGTCTTCTGTGAGCTTC 3840
Db 15652 CGCGACTTCTTGAAGACCAAACTACCGGATGATGCTGCTCAAGTCTTCTGTGAGCTTC 15711
Qy 3841 GATGCTTGGCGGCTGACCTTCAAGCGGAGTTGATGCTTAAGCGCTTGGAGCGGAGAG 3900
Db 15712 GATGCTTGGCGGCTGACCTTCAAGCGGAGTTGATGCTTAAGCGCTTGGAGCGGAGAG 15771
Qy 3901 GATACCTGTCGCGCGGCACTTGGGAGCAGCGGCGGCAAGCGGAGCGCTTGGAGAGATG 3960
Db 15772 GATACCTGTCGCGCGGCACTTGGGAGCAGCGGCGGCAAGCGGAGCGCTTGGAGAGATG 15831
Qy 3961 CTGCTGCGGATGTAAGGAGGTGCTCGGCGTGAAGGTGATCGGCTTCAGCAGAGCTTC 4020
Db 15832 CTGCTGCGGATGTAAGGAGGTGCTCGGCGTGAAGGTGATCGGCTTCAGCAGAGCTTC 15891
Qy 4021 GTGCTTGGGTGGAATTCATTCATGCTTGGCATGAGAGAGCTTGTGCAAGAGG 4080
Db 15892 GTGCTTGGGTGGAATTCATTCATGCTTGGCATGAGAGAGCTTGTGCAAGAGG 15951
Qy 4081 CTGATGAGAGATCGGATCAAGGAGTGTTCAGATCCGGAACCTCGGCTCGCTGCG 4140
Db 15952 CTGATGAGAGATCGGATCAAGGAGTGTTCAGATCCGGAACCTCGGCTCGCTGCG 16011
Qy 4141 TCCGCTTGGCGCGAGACTCGAGAGATCTAGATCAGCGCGGCAATGACAGACCGAGTG 4200
Db 16012 TCCGCTTGGCGCGAGACTCGAGAGATCTAGATCAGCGCGGCAATGACAGACCGAGTG 16071
Qy 4201 GAGGTTCCGCGCAGAGGCGAGAGAGCTTACGCTTAA 4233
Db 16072 GAGGTTCCGCGCAGAGGCGAGAGAGCTTACGCTTAA 16104


```
/db_xref="GI:7453559"
/translation="MEQSSSAIVIGSGRFPGARLDLDFWRMLADGETVORFSEQ
ELASGYDPLVLDVYVYRQAGSVLEVDVRFPAAPFGISPRAEIMDQPHRPMCEAK
ALENAGYDPTAVESIGVYAGAMSSVLTSLHHPMNRMPGFQTLIGNDKXYLAT
HVSRYRLRGPSSIVQTKSTSLVAHLACSLDRECDMAAGITVRLPHRGIVY
AEGGI FSPDGHCRAPDAKANGTIMGNGGVLLRPLRSLDGPVAVLIGSTAND
GARITGTPA BEVGOAQLMEALAGVARSIOY IETHGTLTGDA IETNALRYF
DRDSTRSCAIGSKTGI GHELSAAGIAGIKTVLALHEHRLPSPINFEPSNIDIF
ASSPVYNTSKDMNTGSTPRAGVSSFGTGTVAHVLIEAPAPAKI PAAPASATL
FVYSKSAALDAARLRDLQAHQSLSDVA FSLATTRS PHEHLMAAAREAL
REGIDAAARGOTPPGAVRGCS PGNVRFVFPFGQSQWQMKRQLAEVPHAL
SICDRAI QAEGWSLAEVLAD EGSQIERIDVQPVLFALVAFALMRWGQBPV
VITSGMGEVAAHVAGALSLEDAVAII CRSRILRRI SGQEMVETELSLAEAAAR
GYEDRVAVANSRSTVLSGEPAIGEVLSNAKGVCFRRVVDVASHPOVDPR
EDLALGGLRPRAAVPMRSTVTGAVAGELGANYMNNLRQVPAEVVQALOG
GHGIFVMSBPRLITTSVEEMRBAQAAGAVSLRGQDRPMLIALGAMQGP
VPMGRLPAGGRVPLPTYPWQREYVITAPAKSAGDRRCVRGHPHLDQTLST
QSTRLEWTLIDLRLPMLGDHVRQGA VFFPGAALVLEMAISSAEALDGLQTTDV
LAELAPAGDAVAVOVVTTBEPGRLOFQIASRAPGAGHSAELMRGEELGRVTE
VPALITLSAVARLQASMPAAATYAELETMLQYGPAGQIAELMRGEELGRVTE
DAAGSAEYRLHPALDACFOVGSIFAGGGEATPMVPEVGSRLRLORPSELWCHA
RVNHGROTPROGADFMVVDSGAVVAEUGVYAQRLPGVRREEDDMLEWMP
AAVGTAKVMGRWILLGGGGLGALASMLEAGHVAHAESNTSAAGVALAKAP
DGAQPTAVHLGSLDGGELDPGLAGGALDAPRSADVSPALDPALVRGDSVLTWV
QALAGMGRDAPRLMLTRGAQVAGGVSTQAPLGLGIVIMEDHACVADLD
PTRDDELGALALADDAEVALRGGERCVARIYRQETPRAGRIESCVPDVT
IRADSTYLVNGLGGLGSVAGVLAERAGHVLVGRSGASVORAPALAEARGAR
VTYAKADVADAOERILREYTTSGMPLRGVYHAGILDOGLMQTPARPRKMAK
VQGLHLHALTREAPLSEFVLIVASGVGLSGPGQNTAAATFDLALHHRHAGLPA
LSVDWGLFAEVGMAAQEDRGARLSRGMRSLTPDEGLSALARLGVQVMPVPA
PRLWELVYPAASSRMLSRVLTARASGAGDGLRLREGLAEPPASRLLEPLLR
AOISOVLRLPGKILEVDAPLTSLGMSISMGELRNRIEAMGIVTPATLMTYPTVA
LSGHLAEACGAAVESPHTTADSAVELEESODTLQTLAAKRALT"
16134..37907
/gene="epod"
/length="16134..37907"
/feature="type I polyketide synthase; epothilone PKS modules
3, 4, 5, and 6"
/codon_start=1
/translation="MTTRGPTAONPLKQAAIIIOLEERLGAQAELEBTEPIATV
GICRFPFGADAPAPFWELDAERDAVQEDRRLNLYGVAVEVPHRAGILTEPIATC
FDAFPGISPREASLDPOHRLLEVAEGLDAGIIPRSIDGRTGVFGVAGIADYDIA
RTVARLPREERDASATGNMLSIAAGRLSYTLGQGPCTVDTGSSSLVAIHLACS
LRAGESLDLAGVSTLSPDMEAAARATOLSPDGRCTEDASANGFVGECCGLVY
LKRLSDAORDDRIMALIRGSAINHDSRGTLTAPNVLAQETVIRLRLASHVAGAV
DYVETHTGTSIDGPIEVELRATVGPASDGTCTCAVATVNI GHL EAAAGVGLIK
AALSLTHERIPRINLFRITLIRILBSALALATEPVPPTDPRRAGVSSFMSGT
NAHVLEAPAVELMPAPAPERSAELVLSGSEBALDQAARLEHIDMPELLELGV
AFSLATTRSAPVTHRLAAVTSREGLALASVAGQTPAGARLAIASSRKLALGT
GOGATQEMGRGLCAAMPAPFEADRCVTLTDFRDLRPLAEVMAAESGSLIDDOT
AFTQPALFAVEYALTLAMRSWGVPELIVGELVLAACVAVESLIDEGRIYVAAAG
RLMGSLAGGAMVSLGAPAEAVAAVA PHAMVSI IAVNGEYVIVAGVEA VQALTA
```

Query Match 97.2%; Score 4114.6; DB 1; Length 58733;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 4159; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
181 ATGTCGCCGCCGCCGAGCGGCA CGCTCCGTTTCTCTCA CAGACATCCAGAATCC 240
6440 ATCTGTCGCCGCCGCCGAGCGGCA CGTTCGTTTCTCTCA CAGACATCCAGAATCC 6499
241 TACTGCTGGGCGCGGACAGGACGCTTTACGCTCCGACGGGATCCAGCCTATCGGAA 300
6500 TACTGCTGGGCGCGGACAGGACGCTTTACGCTCCGACGGGATCCAGCCTATCGGAA 6559
301 TACGACTGTACGATCTTCGACGTGCGGACGCTGAGCCGCCCTTTCCGAAAGTGTGGCG 360
6560 TACGACTGTACGATCTTCGACGTGCGGACGCTGAGCCGCCCTTTCCGAAAGTGTGGCG 6619
361 CGGACACACATGCTTCCGGGCGCAACAGCCTGCCGACATGATGACAGTCACTTAA 420
6620 CGGACACACATGCTTCCGGGCGCAACAGCCTGCCGACATGATGACAGTCACTTAA 6679
421 GTCGACGCCGACATCGAGATCATGATCTGCGCGGGCTCGACCGGACACACGGGAAAGCG 480
6680 GTCGACGCCGACATCGAGATCATGATCTGCGCGGGCTCGACCGGACACACGGGAAAGCG 6739
481 AGGCTCGTGTCTGGCGGACATGCGATGTCGACCGGATCTTATGACACCGAGCCCTCCG 540
6740 AGGCTCGTGTCTGGCGGACATGCGATGTCGACCGGATCTTATGACACCGAGCCCTCCG 6799
541 CTCTACACGCTCGTCCGCGCTTCCGCTGACGAGCGGAAACCGCTCTGTCGATGATC 600
6800 CTCTACACGCTCGTCCGCGCTTCCGCTGACGAGCGGAAACCGCTCTGTCGATGATC 6859
601 GATCTCATTAAGCTTGAACCTAGGACGCTGTCATCTTCAAGACCTGCTCAGCTTC 660
6860 GATCTCATTAAGCTTGAACCTAGGACGCTGTCATCTTCAAGACCTGCTCAGCTTC 6919
661 TACGAAATCCCGAGACCTCTCTCTCTGTCGATGTCGATGTCGATGTCGATGTC 720
6920 TACGAAATCCCGAGACCTCTCTCTCTGTCGATGTCGATGTCGATGTCGATGTC 6979
721 GCGCTGAGTCTCGCAAGAGTCTGAGGCGATGAAGCGATGAAGTTCGGAAGCGG 780
6980 GCGCTGAGTCTCGCAAGAGTCTGAGGCGATGAAGCGATGAAGTTCGGAAGCGG 7039
781 CGCATTCGCGAGCTCCACCTCCGCGGACGCTTCGATGAAAGCGGATTCATCCCTG 840
7040 CGCATTCGCGAGCTCCACCTCCGCGGACGCTTCGATGAAAGCGGATTCATCCCTG 7099
841 AAGGAGATCCGCTTCCGCGACACGAGACATAGCTGCGCTGGAATCTCTGGGGTCAATTG 900
7100 AAGGAGATCCGCTTCCGCGACACGAGACATAGCTGCGCTGGAATCTCTGGGGTCAATTG 7159
901 AAGCGGCGTGTCCGGGAGCGCGGGCTGACCCCGACGGGCGTCATCTGCGTGAATTTCC 960
7160 AAGCGGCGTGTCCGGGAGCGCGGGCTGACCCCGACGGGCGTCATCTGCGTGAATTTCC 7219
961 GAGGTGATCGGGCGCTGAGACGCGAGCCCGGTTTACGCTCAACATAGCTTTCAC 1020
7220 GAGGTGATCGGGCGCTGAGACGCGAGCCCGGTTTACGCTCAACATAGCTTTCAC 7279
1021 CGGCTCCCGCTCCATCCGCGCTGTAACGATATCACCGGGACTTCACTGATGATGCTCTC 1080
7280 CGGCTCCCGCTCCATCCGCGCTGTAACGATATCACCGGGACTTCACTGATGATGCTCTC 7339
1081 CTGGAATCGACACCACTCGCGACCAAGAGCTTGAACAGGCGGTTAGCGTATTCAGAG 1140
7340 CTGGAATCGACACCACTCGCGACCAAGAGCTTGAACAGGCGGTTAGCGTATTCAGAG 7399
1141 CAGCTGTGGAGAGGATGATCACTGCGAGTAAAGCGGTTACAGGTTCAAGGAGGCC 1200
7400 CAGCTGTGGAGAGGATGATCACTGCGAGTAAAGCGGTTACAGGTTCAAGGAGGCC 7459
1201 GCGCGGCTCTGGGAGATCCACGAGGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTT 1260
7460 GCGCGGCTCTGGGAGATCCACGAGGCGCATTTGTTCCCGTGTGCTCAAGAGCGGCTC 7519
```

QY	1261	AACCAAGCAAGTCGTTGGTGTCACTCGTTGCAGAGGCTCCGGAATCCCGGTGAACAACAGC	1320
Db	7520	AACCAAGCAAGTCGTTGGTGTCACTCGGTGAGAGGCTCCGGAATCCCGGTGAACAACAGC	7579
QY	1321	ACGCAAGACTCCCTCAAGCTGCTGTGGATCATCAGAGCTCTACAGACGACAGATAGGGGAACTCGTC	1380
Db	7580	ACGCAAGACTCCCTCAAGCTGCTGTGGATCATCAGAGCTCTACAGACGACAGATAGGGGAACTCGTC	7639
QY	1381	CTCGCGTGGGACATCTGTGACGAGAGTGTTCCCGCCCGACCTTCTGAGCAGCATCTCGAA	1440
Db	7640	CTCGCGTGGGACATCTGTGACGAGAGTGTTCCCGCCCGACCTTCTGAGCAGCATCTCGAA	7699
QY	1441	GCCTAACGTCGTTTTTCTCCGGCGGCTCACTGAGGAAACAAGGGGTGAACAAGTGCCTGT	1500
Db	7700	GCCTAACGTCGTTTTTCTCCGGCGGCTCACTGAGGAAACAATGAGTGAACAATGCGCTGT	7759
QY	1501	TCGCTTCCGCGCTGCGCCAGCTGAGAAAGCGGGGCGAGCGGAAACGCGAACGAGCGCTGCTG	1560
Db	7760	TCGCTTCCGCGCTGCGCCAGCTGAGAAAGCGGGGCGAGCGGAAACGAGAACCAATCGCTGCTG	7819
QY	1561	AGCGAGCAATACGCTGACAGGCGCTGTTCCGCGCGCGGGGTGAGACAGCTGCCATGCAAGCTC	1620
Db	7820	AGCGAGCAATACGCTGACAGGCGCTGTTCCGCGCGCGGGGTGAGAGAGCTGCTATGCAAGCTC	7879
QY	1621	GCCGTGCTGTGGGCGCGCGGACGACGTCACGTAAGAAAGCTTTGGCGCGCTTCCGCGGGA	1680
Db	7880	GCCGTGCTGTGGGCGCGCGGACGACGTCACGTAAGAAAGCTTTGGCGCGCTTCCGCGGGA	7939
QY	1681	CTTGGCGCGCGGCGCTGCGGAGACAGGGGGGACAGCCCGGAACAATTTGATGATGCGGTGATG	1740
Db	7940	CTTGGCGCGCGGCGCTGCGGAGACAGGGGGGACAGCCCGGAACAATTTGATGATGCGGTGATG	7999
QY	1741	GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCTCAC	1800
Db	8000	GAGAAAGGCTGGGAGCAGGTTGTCGCGGTTCTCGCGGTGCTCGAGTCAGGCGCGGCTCAC	8059
QY	1801	GTGCCGATCGATGCCGACCTACACGGCGGAGCGTATCCACTACCTCTCGATCATGTGAG	1860
Db	8060	GTGCCGATCGATGCCGACCTACACGGCGGAGGTAATCCACTACCTCTCGATCATGTGAG	8119
QY	1861	GTAAGAGCTCGAGCTGAGCGCAGCAATGCTGGAATGGCAACTGATAGGCGCGCGGGATC	1920
Db	8120	GTAAGAGCTCGAGCTGAGCGCAGCAATGCTGGAATGGCAACTGATAGGCGCGCGGGATC	8179
QY	1921	CAGCGGCTGTCTGTGAGCGAGGCGCGGCTCGAAAGCGACGCGGACCAAGCTCCGATGATG	1980
Db	8180	CAGCGGCTGTCTGTGAGCGAGGCGCGGCTCGAAAGCGACGCGGACCAAGCTCCGATGATG	8239
QY	1981	CCCATTCAGACACCTTCCGATCTCGGATGTGCGCTATCATCTACCTCGGGATCCACAGGGTTG	2040
Db	8240	CCCATTCAGACACCTTCCGATCTCGGATGTGCGCTATCATCTACCTCGGGATCCACAGGGTTG	8299
QY	2041	CCCAAGGGGGGATGATGATCATTCGGGGGTGCGGTCAACACATCTGAGCATCAACGAG	2100
Db	8300	CCCAAGGGGGGATGATGATCATTCGGGGGTGCGGTCAACACATCTGAGCATCAACGAG	8359
QY	2101	CGCTTCGAATAAGGCGCGGAGACAGGAGTGTGAGCGCTCTCGCTGAGCTTGCATCTC	2160
Db	8360	CGCTTCGAATAAGGCGCGGAGACAGGAGTGTGAGCGCTCTCTCGCTGAGCTTGCATCTC	8419
QY	2161	TCGGTCTATGATGTGTTTCGGGATCTCTGGCGCGGGCGGTGAGATCTGTGTCGGACCG	2220
Db	8420	TCGGTCTATGATGTGTTTCGGGATCTCTGGCGCGGGCGGTGAGATCTGTGTCGGACCG	8479
QY	2221	TCCAAGCTGCGCGATCCGGCGCATTTGGGCAAGTTATCCGAACGAGAAAGTGAACGCTG	2280
Db	8480	TCCAAGCTGCGCGATCCGGCGCATTTGGGCAAGTTATCCGAACGAGAAAGTGAACGCTG	8539
QY	2281	TGGAAGCTCGGTGCGCGGCTGATGTGGGATGCTCGTCAGACATTTTGAAGGTGCGCCGAT	2340
Db	8540	TGGAAGCTCGGTGCGCGGCTGATGTGGGATGCTCGTCAGACATTTTGAAGGTGCGCCGAT	8599
QY	2341	TCGCTCGCTAGATCTTCGCGGCTTTCGCTGCTGAGCGGACGCACTGAGTCCGATGGGCTG	2400

Db	8600	TCGCTCCCTAGAGTCTTCGCGACTTTCGCTCTGTAGCGGCACTGAGATCCCGTGGGCTCG	8659
Qy	2401	CTTGGCGAGCTCCAGGCAATCAGGCCCGGGGTGTGGTGAATCAGCTTGGGCGGGGCCACC	2460
Db	8660	CTGGCGAGCTCCAGGCAATCAGGCCCGGGGTGTGGTGAATCAGCTTGGGCGGGGCCACC	8719
Qy	2461	GAAGCGTGCATCTGTGCATTCGGGTATCCCGTGAAGAAAGTCCGACTTATCTGTGGCGAGC	2520
Db	8720	GAAGCGTGCATCTGTGCATTCGGGTATCCCGTGAAGAAAGTCCGACTTATCTGTGGCGAGC	8779
Qy	2521	ATCCCTTAACGGCCGTCCGCTGGCCGCAACAGAGTTCCAGTGGCTCATGTAGAGGGCTTGAA	2580
Db	8780	ATCCCTTAACGGCCGTCCGCTGGCCGCAACAGAGTTCCAGTGGCTCATGTAGAGGGCTTGAA	8839
Qy	2581	CCGCGCCCGGCTTGGGTTCGGGGCAACTTACATTGGCGGGGTGGGCTTGGCACTTGGGC	2640
Db	8840	CCGCGCCCGGCTTGGGTTCGGGGCAACTTACATTGGCGGGGTGGGCTTGGCACTTGGGC	8899
Qy	2641	TACTGGCGCGATGAAGAAAGACGCGCAAGACTTCTCGTGCACTCCCGAGACGGGGAG	2700
Db	8900	TACTGGCGCGATGAAGAAAGACGCGCAAGACTTCTCGTGCACTCCCGAGACGGGGAG	8959
Qy	2701	CGCCTTACAAAGACCGGGGATCTGGGGCGGATCCTGCGCATGGAAATCATCGATTATG	2760
Db	8960	CGCCTTACAAAGACCGGGGATCTGGGGCGGATCCTGCGCATGGAAATCATCGATTATG	9019
Qy	2761	GGGCGTAGAGCAACAATAACAAGCTTTCGGCGATATCCGCTTGAAGCTTCGGGGAAATCGAG	2820
Db	9020	GGGCGTAGAGCAACAATAACAAGCTTTCGGCGATATCCGCTTGAAGCTTCGGGGAAATCGAG	9079
Qy	2821	GAAGCGCTCAAGTGCATCCGAAAGTACGCGACGCGGTGATTTGTCCCTTCGGGAACGAC	2880
Db	9080	GAAGCGCTCAAGTGCATCCGAAAGTACGCGACGCGGTGATTTGTCCCTTCGGGAACGAC	9139
Qy	2881	GGCGCGAAACAAGCTCTTTCAGCTTATGTGTCTCCCGAAGGGCACACGGGAAACCGCTGGC	2940
Db	9140	GGCGCGAAACAAGCTCTTTCAGCTTATGTGTCTCCCGAAGGGCACACGGGAAACCGCTGGC	9199
Qy	2941	GAGCAGAGACGGAGCCTCAAGACCCGAGCGGATCGACGCGAGAGCAACGCGCGCGAAGCG	3000
Db	9200	GAGCAGAGACGGAGCCTCAAGACCCGAGCGGATCGACGCGAGAGCAACGCGCGCGAAGCG	9259
Qy	3001	GACGCGTTGAACGACGCGCGAGAGGGGTGAGTTCAAGCTGCTCGACACGGAATTCGGAGG	3060
Db	9260	GACGCGTTGAACGACGCGCGAGAGGGGTGAGTTCAAGCTGCTCGACACGGAATTCGGAGG	9319
Qy	3061	GACCTGGAACGGAAGCCCGTCTGTGATCTTACCGGGAGAGATCCGCGGAGAGGGGGGCTG	3120
Db	9320	GATTTGACGGAAGCCCGTCTGTGATCTTACCGGGAGAGATCCGCGGAGAGGGGGGCTG	9379
Qy	3121	GACGCTTACGCGCGCTGCGCTGAGAGGTTCGAAAGCTTCTTAGAGCCCGGATTCGCTTGT	3180
Db	9380	GACGCTTACGCGCGCTGCGCTGAGAGGTTCGAAAGCTTCTTAGAGCCCGGATTCGCTTGT	9439
Qy	3181	GAGTTTGTGCGATTCTGAGCTGTGTAGGACGCTGGAGCCCGACGGCGGACCCCTTCCC	3240
Db	9440	GAAATTCGCGCGATTCTGAGCTGTGTAGGACGCTGGAGCCCGACGGCGGACCCCTTCCC	9499
Qy	3241	AAATTCCTGTTATCTATGGCGGGGAGACGTAACCCCGTTCGAAACCTACGGGTATGTCAA	3300
Db	9500	AAATTCCTGTTATCTATGGCGGGGAGACGTAACCCCGTTCGAAACCTACGGGTATGTCAA	9559
Qy	3301	TCCGCGCGCATCGAGGGCGTGGAGCAAGAGGCTTCTATTAATTAACAACCGTTTCGAGACCGT	3360
Db	9560	TCCGCGCGCATCGAGGGCGTGGAGCAAGAGGCTTCTATTAATTAACAACCGTTTCGAGACCGT	9619
Qy	3361	TTGCTGAAGCTCTTCGATCAACGGGATCGAGCGCGGACGCAAGTTTCGCAAAACTTTCGAC	3420
Db	9620	TTGCTGAAGCTCTTCGATCAACGGGATCGAGCGCGGACGCAAGTTTCGCAAAACTTTCGAC	9679
Qy	3421	GTGTTTCAATGAGCGGCGTTCAACTTCTGTGTGTGGGCAAGATCGACGCTATGAGTCC	3480

Db 9680 GTGTCATGAAACGCGCTCGCCTCCTGTTCTGGGAGAGATGATCCATCGAGTCG 9739
Qy 3481 CTGATGAGATGCTGCTCGCGAGAAATTTTGCTGTGAGGCGCGATATATGCGAGCTC 3540
Db 9740 CTGATGAGATGCTGCTCGCGAGAAATTTTGCTGTGAGGCGCGATATATGCGAGCTC 9799
Qy 3541 CTGATGAGAGGCGCTTCTGCAACATCGGCGTCTGTCGGTGGGCAATTCAATTTT 3600
Db 9800 CTGATGAGAGGCGCTTCTGCAACATCGGCGTCTGTCGGTGGGCAATTCAATTTT 9859
Qy 3601 GAAACAGTTGCGCGCGTCTGCAACCTGAGACATTCGAGCTTACGTACAGAGGATGCTG 3660
Db 9860 GAAACAGTTGCGCGCGTCTGCAACCTGAGACATTCGAGCTTACGTACAGAGGATGCTG 9919
Qy 3661 GCGCGCGCGGATGAGACCGCGCGAGTTCCAGGCTGTGAGCTCGATCAGGATTTCTCAACG 3720
Db 9920 GCGCGCGCGGATGAGACCGCGCGAGTTCCAGGCTGTGAGCTCGATCAGGATTTCTCAACG 9979
Qy 3721 AGCGCGCGCAAGACGCGCGCGCTTCCCGCGCGAGACACTTCCCGATATGCTT 3780
Db 9980 AGCGCGCGCAAGACGCGCGCGCTTCCCGCGCGAGACACTTCCCGATATGCTT 10039
Qy 3781 CGGCACTTCTTGAGAGCAAACTAACCCGAGTACATGATGCTTACGATCTTGTGAGCTC 3840
Db 10040 CGGCACTTCTTGAGAGCAAACTAACCCGAGTACATGATGCTTACGATCTTGTGAGCTC 10099
Qy 3841 GATCGCTTCCGCTGACCTCCAAACGAGCAAGGTCATCTGTAAGGCTTCCGCGAGCGAAG 3900
Db 10100 GATCGCTTCCGCTGACCTCCAAACGAGCAAGGTCATCTGTAAGGCTTCCGCGAGCGAAG 10159
Qy 3901 GATACCTGTCGCGCGCGCATTTGGGGGCAACGCGCGCAAGCGCTTGGAGAGATC 3960
Db 10160 GATACCTGTCGCGCGCGCATTTGGGGGCAACGCGCGCAAGCGCTTGGAGAGATC 10219
Qy 3961 CTCGTCGCGGTCGTAACGAGGAGTGTCTCGGCTGAGAGTGTCTCGGCTTCCAGAGAGCTTC 4020
Db 10220 CTCGTCGCGGTCGTAACGAGGAGTGTCTCGGCTGAGAGTGTCTCGGCTTCCAGAGAGCTTC 10279
Qy 4021 GTGCATCTTGTGCGACATCATCATCTGTCGATGAGAGAGCTTGTCCAGAGAGG 4080
Db 10280 GTGCATCTTGTGCGACATCATCATCTGTCGATGAGAGAGCTTGTTCAGAGAGG 10339
Qy 4081 CTGATGAGAGATGCGCATCAACGAGTGTTCAGTACCGAACTTCGGCTCGTGGCG 4140
Db 10340 CTGATGAGAGATGCGCATCAACGAGTGTTCAGTACCGAACTTCGGCTCGTGGCG 10399
Qy 4141 TCCGCTTTCGCGCGAGATCTGAGAGATCTAGATCAGCGCGCGAATCAGAGCCGAGTG 4200
Db 10400 TCCGCTTTCGCGCGAGATCTGAGAGATCTAGATCAGCGCGCGAATCAGAGCCGAGTG 10459
Qy 4201 GAGGTTGCGCGCAAGGCGAGAGAGCTGAGCTTA 4233
Db 10460 GAGGCTGCGCGCAAGGCGAGAGAGCTGAGCTTA 10492

RESULT 9
ARI172664
LOCUS ARI172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION ARI172664
VERSION ARI172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epoithlones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES location/Qualifiers
SOURCE 1..71989
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others

ORIGIN
Query Match 97.2%; Score 4114.6; DB 6; Length 71989;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 4159; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

Qy 1 ATGACATCAATGACTTCTGAAAGAGCTCGAGCAAGGATATCAAGCTGGCGCGAT 60.
Db 6260 ATGACATCAATGACTTCTGAAAGAGCTCGAGCAAGGATATCAAGCTGGCGCGAT 6319
Qy 61 GGGAGAGCGCTTCCAGATACAGGCGCCCAAGAAAGCCCTTGAACCCGAACTCTGCTCGA 120
Db 6320 GGGAGAGCGCTTCCAGATACAGGCGCCCAAGAAAGCCCTTGAACCCGAACTCTGCTCGA 6379
Qy 121 ATTCGAGCACAAAAGACATCTGAGATGCTCTCGTGAGAGCTCCCGAGAAATCC 180
Db 6380 ATTCGAGCACAAAAGACATCTGAGATGCTCTCGTGAGAGCTCCCGAGAAATCC 6439
Qy 181 ATCTGCGCGCGCCAGCGAGCGGACGCTCGTTTCTCTCAGAGATCCAAAGATCC 240
Db 6440 ATCTGCGCGCGCCAGCGAGCGGACGCTCGTTTCTCTCAGAGATCCAAAGATCC 6499
Qy 241 TACTGCGTGGCGCGAGACGAGGCTTTAAGCTCCCGAGGAGATCCAGCTTATCGGAA 300
Db 6500 TACTGCGTGGCGTGGAGACGAGGCTTTAAGCTCCCGAGGAGATCCAGCTTATCGGAA 6559
Qy 301 TACGACTGTACGATCTCGAGCTGCGAGGCTGAGCGCGCTTTCGGAAGGTGCGG 360
Db 6560 TACGACTGTACGATCTCGAGCTGCGAGGCTGAGCGCGCTTTCGGAAGGTGCGG 6619
Qy 361 CGGACAGCATGCTTTCGGGCGCAACAGCTGCGCCGACATGATGAGTATGAGACTTAA 420
Db 6620 CGGACAGCATGCTTTCGGGCGCAACAGCTGCGCCGACATGATGAGTATGAGACTTAA 6679
Qy 421 GTCAAGCGCGACATCGATCATGATCTGCGCGGCTCCAGCGGAGACACGCGAAGCG 480
Db 6680 GTCAAGCGCGACATCGATCATGATCTGCGCGGCTCCAGCGGAGACACGCGAAGCG 6739
Qy 481 AGGCTGCTGTGCGAGATGAGATGCGAACCGCATCTTATGACACCGAGCGGCTCCG 540
Db 6740 AGGCTGCTGTGCGAGATGAGATGCGAACCGCATCTTATGACACCGAGCGGCTCCG 6799
Qy 541 CTCTATACGTCGTCGCGCTTTCGCGCTGAGAGAGCGGCAACCCGCTCTGCTCAGATC 600
Db 6800 CTCTATACGTCGTCGCGCTTTCGCGCTGAGAGAGCGGCAACCCGCTCTGCTCAGATC 6859
Qy 601 GATCTCATTAACGTTGACTGAGAGCTGCTTCATCATCTTCAAGAGACTGGCTAGCTTC 660
Db 6860 GATCTCATTAACGTTGACTGAGAGCTGCTTCATCATCTTCAAGAGACTGGCTAGCTTC 6919
Qy 661 TACGAATCCCGAGACCTCTCCCTGCTCTGAGAGCTCTGTAACCGGATATGATCTC 720
Db 6920 TACGAATCCCGAGACCTCTCTCCCTGCTCTGAGAGCTCTGTAACCGGATATGATCTC 6979
Qy 721 GCGCTGAGTCTCGCAAGAGTCTGAGCGGCATCAAGATGATGATTAAGTGAAGCG 780
Db 6980 GCGCTGAGTCTCGCAAGAGTCTGAGCGGCATCAAGATGATGATTAAGTGAAGCG 7039
Qy 781 CGCATCGCGAGCTCCCACTTCGCGAGCGCTTCCGATGAGAGCGCATTCATCCCTG 840
Db 7040 CGCATCGCGAGCTCCCACTTCGCGAGCGCTTCCGATGAGAGCGCATTCATCCCTG 7099
Qy 841 AAGGAGATCCGCTTCGCGACACGAGCAATGCTGCGCTCGACTCTTGAGAGTGAATG 900
Db 7100 AAGGAGATCCGCTTCGCGACACGAGCAATGCTGCGCTCGACTCTTGAGAGTGAATG 7159
Qy 901 AAGCGCGTTCGCGAGAGCGCGGCTGAGACCCGAGCGGCTCATCTCGCTGATTTTC 960
Db 7160 AAGCGCGTTCGCGAGAGCGCGGCTGAGACCCGAGCGGCTCATCTCGCTGATTTTC 7219
Qy 961 GAGGTATCGGCGCTGAGAGCGGAGCGGCTTTCGCTCAACATCAAGCTCTTCAAC 1020
Db 7220 GAGGTATCGGCGCTGAGAGCGGAGCGGCTTTCGCTCAACATCAAGCTCTTCAAC 7279

QY 1021 CCGCTCCCCGTCATCCGCGCTGAAAGATATCAACCGGGACTTCAACGTCATGTCCTC 1080
DB CCGCTCCCCGTCATCCGCGCTGAAAGATATCAACCGGGACTTCAACGTCATGTCCTC 7339
QY 1081 CTGGAATGACCACTCGCGCAAGAGACTTCAAGACGCGCTTAAGGCTATTCAAG 1140
DB CTGGAATGACCACTCGCGCAAGAGACTTCAAGACGCGCTTAAGGCTATTCAAG 7399
QY 1141 CAGCTGTGGAAAGCATGATGATCACTGCAAGTAAAGGCTATTCAGGTCAGGAGAGCC 1200
DB CAGCTGTGGAAAGCATGATGATCACTGCAAGTAAAGGCTATTCAGGTCAGGAGAGCC 7459
QY 1201 GCCCGGCTCTGGGATCAAGAGGCGGATTTCCCGTGTGTCAAGAGCGGCTT 1260
DB GCCCGGCTCTGGGATCAAGAGGCGGATTTCCCGTGTGTCAAGAGCGGCTT 7519
QY 1261 AACCAAGATGATGATGATCACTGTCAGAGAGGCTGAAACTCCGCTGTACACAGC 1320
DB AACCAAGATGATGATGATCACTGTCAGAGAGGCTGAAACTCCGCTGTACACAGC 7579
QY 1321 AGCAGACTCTCAGCTGTCTGATCATCAGCTTCAAGACAGATGGAGACCTCTCTC 1380
DB AGCAGACTCTCAGCTGTCTGATCATCAGCTTCAAGACAGATGGAGACCTCTCTC 7639
QY 1381 CTCGCTGGGACATGTCAGAGAGTTCCTCCCGGACCTTCTGAGACATCTCTCGA 1440
DB CTCGCTGGGACATGTCAGAGAGTTCCTCCCGGACCTTCTGAGACATCTCTCGA 7699
QY 1441 GCGTACGTGTTTCTCCGCGGCTCATCTGAGAAACATGGAGTGAACAGGTCGCTCT 1500
DB GCGTACGTGTTTCTCCGCGGCTCATCTGAGAAACATGGAGTGAACAGGTCGCTCT 7759
QY 1501 TCGCTTCGCTGCCAGCTAAGAGCGGCGAGCAACGCAACGCAACGCGCTCTG 1560
DB TCGCTTCGCTGCCAGCTAAGAGCGGCGAGCAACGCAACGCAACGCGCTCTG 7819
QY 1561 AGCGAGCATACGCTGACGCGCTTTCGCGCGCGGCTGACAGTCCCATGACCTC 1620
DB AGCGAGCATACGCTGACGCGCTTTCGCGCGCGGCTGACAGTCCCATGACCTC 7879
QY 1621 GCCGTGATGTCGCGCGCAAGAGCTCATGTAAGAGCTTTCGCGCGCTTTCGCGCG 1680
DB GCCGTGATGTCGCGCGCAAGAGCTCATGTAAGAGCTTTCGCGCGCTTTCGCGCG 7939
QY 1681 CTTGGCGCGGCTGCGAGAGAGGCGGAGCAGCCCGAACATTTGTCGCGTGTATG 1740
DB CTTGGCGCGGCTGCGAGAGAGGCGGAGCAGCCCGAACATTTGTCGCGTGTATG 7999
QY 1741 GAGAAAGCTGGAGAGAGTTCGCGGTTTCGCGGCTGACAGTCAAGCGCGCTAC 1800
DB GAGAAAGCTGGAGAGAGTTCGCGGTTTCGCGGCTGACAGTCAAGCGCGCTAC 8059
QY 1801 GTGCCGATGATGCGGACTTACCGGCGAGGCTATCATCTCTCTGATGATGATG 1860
DB GTGCCGATGATGCGGACTTACCGGCGAGGCTATCATCTCTCTGATGATGATG 8119
QY 1861 GTAAAGCTGTGCTGACGAGCATGATGAGCAACTGTCTCTCTGATGATGATG 1920
DB GTAAAGCTGTGCTGACGAGCATGATGAGCAACTGTCTCTCTGATGATGATG 8179
QY 1921 CAGCGGCTGCTGAGAGAGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB CAGCGGCTGCTGAGAGAGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8239
QY 1981 CCGATTGACACCTTCCGATCTCGGATGATCATCTACACCTCCGAGTCAAGGGTTG 2040
DB CCGATTGACACCTTCCGATCTCGGATGATCATCTACACCTCCGAGTCAAGGGTTG 8299
QY 2041 CCGAAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100
DB CCGAAGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8359

QY 2101 CCGTTGAAATTAAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
DB CCGTTGAAATTAAGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8419
QY 2161 TCGGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
DB TCGGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8479
QY 2221 TCCAGCTCCGCGATCCGCGGCTTGGGCAAGAGTTCAGAACAGAGAGAGAGAGAG 2280
DB TCCAGCTCCGCGATCCGCGGCTTGGGCAAGAGTTCAGAACAGAGAGAGAGAGAG 8539
QY 2281 TGAATGCGGAGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 2340
DB TGAATGCGGAGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 8599
QY 2341 TCGCTGCTAAGCTCTGCGGCTTTCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
DB TCGCTGCTAAGCTCTGCGGCTTTCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 8659
QY 2401 CCGGCGAGCTTCAGAGCATCAGGCGCGGCTGATGATGATGATGATGATGATGATGAT 2460
DB CCGGCGAGCTTCAGAGCATCAGGCGCGGCTGATGATGATGATGATGATGATGATGAT 8719
QY 2461 GAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
DB GAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8779
QY 2521 ATCCCTTAAGGCGCGCTGCGGCTGCGAACAGAGTTCAGAGTTCAGAGAGAGAGAG 2580
DB ATCCCTTAAGGCGCGCTGCGGCTGCGAACAGAGTTCAGAGTTCAGAGAGAGAGAG 8839
QY 2581 CCGGCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCT 2640
DB CCGGCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCTTTCGCGGCT 8899
QY 2641 TACTGCGGATGAG 2700
DB TACTGCGGATGAG 8959
QY 2701 CCGCTCTAAGAGCGGCGATCTGGGCGGCTACCTGCGGAGAGAGAGAGAGAGAGAG 2760
DB CCGCTCTAAGAGCGGCGATCTGGGCGGCTACCTGCGGAGAGAGAGAGAGAGAGAG 9019
QY 2761 GGGCGTGAAGACCAATCAAGTTCGCGGATTCAGGCTTTCGCGGATTCGCGGATTC 2820
DB GGGCGTGAAGACCAATCAAGTTCGCGGATTCAGGCTTTCGCGGATTCGCGGATTC 9079
QY 2821 GAAAGCTCAAGTGCATCCGAGAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
DB GAAAGCTCAAGTGCATCCGAGAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9139
QY 2881 GCGGAGAACAGGCTCTTCTAGCTTGTGTCGCGAGAGAGAGAGAGAGAGAGAGAG 2940
DB GCGGAGAACAGGCTCTTCTAGCTTGTGTCGCGAGAGAGAGAGAGAGAGAGAGAGAG 9199
QY 2941 GAGCAGAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
DB GAGCAGAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 9259
QY 3001 GAGGCTTGAAG 3060
DB GAGGCTTGAAG 9319
QY 3061 GACCTGAG 3120
DB GACCTGAG 9379
QY 3121 GACGCTTACAG 3180
DB GACGCTTACAG 9439
QY 3181 GAGTTTGTGATCTCTGAGCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240

Db 9440 GAATTCGGCCCATCTGAGCTGCTGAGCAGCGGAGCCGACGCGCGCCCTTCC 9499
 Qy 3241 AAATTCGTTATCCATGCGCGGCGGAGCAGCAGTACCCCGGCAAACTAGCGCTATGCA 3300
 Db 9500 AAATTCGTTATCCATGCGCGGCGGAGCAGCAGTACCCCGGCAAACTAGCGCTATGCA 9559
 Qy 3301 TCCGCGCGCATGAGAGCGTGGAGCAGAGGCTTCTATTTATTAACACCCGTTGAGACCGT 3360
 Db 9560 TCCGCGCGCATGAGAGCGTGGAGCAGAGGCTTCTATTTATTAACACCCGTTGAGACCGT 9619
 Qy 3361 TTGCTGAAGCTCTCCGATCAGCGGATGAGCGCGGAGCGCAGTTCCGCAAACTTGCAC 3420
 Db 9620 TTGCTGAAGCTCTCCGATCAGCGGATGAGCGCGGAGCGCAGTTCCGCAAACTTGCAC 9679
 Qy 3421 GTGTTCGATGAAGGCGGTTCAACCTCTGTTGCGGCGAGATGAGCGCATGAGTCG 3480
 Db 9680 GTGTTCGATGAAGGCGGTTGCGCTCTGTTGCGGCGAGATGAGCGCATGAGTCG 9739
 Qy 3481 CTGTATGAGATCGTCTGCGAGAAATTTGCTGCTGAGGCGCGGATATATGCGCAGCTC 3540
 Db 9740 CTGTATGAGATCGTCTGCGAGAAATTTGCTGCTGAGGCGCGGATATATGCGCAGCTC 9799
 Qy 3541 CTGATGAGCAGGCGCTTCTGCAACATCGCGCTCTGCTCGGTGGGCAATTCATTTT 3600
 Db 9800 CTGATGAGCAGGCGCTTCTGCAACATCGCGCTCTGCTCGGTGGGCAATTCATTTT 9859
 Qy 3601 GAACAGGTTGCGCGGCTTCTGCAACCTGAGATTCGAGCGTTTACGTGACAGCATGCTG 3660
 Db 9860 GAACAGGTTGCGCGGCTTCTGCAACCTGAGATTCGAGCGTTTACGTGACAGCATGCTG 9919
 Qy 3661 GCGCGCGGAGTGAACCCGCGGCAATTCAGAGTCTGTAAGCTCGGTGAGGATTCCTCACCG 3720
 Db 9920 GCGCGCGGAGTGAACCCGCGGCAATTCAGAGTCTGTAAGCTCGGTGAGGATTCCTCACCG 9979
 Qy 3721 AGCGCGGCGAGCAGCGCGGCGCTTCCGCGCGGAGCAGCATTCGCGGATATGCTT 3780
 Db 9980 AGCGCGGCGAGCAGCGCGGCGCTTCCGCGCGGAGCAGCATTCGCGGATATGCTT 10039
 Qy 3781 CGGCACTTCTGAGAGCAAACTACCCGAGTACATGAGTCTTCTGAGAGTCT 3840
 Db 10040 CGGCACTTCTGAGAGCAAACTACCCGAGTACATGAGTCTTCTGAGAGTCT 10099
 Qy 3841 GATCGGTTGCGGTCGATCCAAACGCGCAAGGTGATGATGAGGCTTCCGCGGAGGAG 3900
 Db 10100 GATCGGTTGCGGTCGATCCAAACGCGCAAGGTGATGATGAGGCTTCCGCGGAGGAG 10159
 Qy 3901 GATACCTGTCGCGCGGATTCGCGGAGCAGCGCGGAGCGCTTGGAGAGATC 3960
 Db 10160 GATACCTGTCGCGCGGATTCGCGGAGCAGCGCGGAGCGCTTGGAGAGATC 10219
 Qy 3961 CTGCTCGGCTCTGACGAGAGTGTCTCGGCTGAGAGTGTCTCGGCTTCCAGAGAGCTTC 4020
 Db 10220 CTGCTCGGCTCTGACGAGAGTGTCTCGGCTGAGAGTGTCTCGGCTTCCAGAGAGCTTC 10279
 Qy 4021 GTGATCTTGTGAGCAGATTCATCGTCTGATGAGAGGCTTGTGAGAGAGG 4080
 Db 10280 GTGATCTTGTGAGCAGATTCATCGTCTGATGAGAGGCTTGTGAGAGAGG 10339
 Qy 4081 CTGATGAGAGATTCGATCAGAGGATTCAGTACCCGAACTTCGAGCTCGCTGCGG 4140
 Db 10340 CTGATGAGAGATTCGATCAGAGGATTCAGTACCCGAACTTCGAGCTCGCTGCGG 10399
 Qy 4141 TCCGCTTGGCGCGAGATTCAGAGATTCATGAGCGCGCGGAACTTCAGAGAGCGAGTG 4200
 Db 10400 TCCGCTTGGCGCGAGATTCAGAGATTCATGAGCGCGCGGAACTTCAGAGAGCGAGTG 10459
 Qy 4201 GAGGTTGCGCGAGAGGAG 4233
 Db 10460 GAGGTTGCGCGAGAGGAG 10492

RESULT 10

AF188287
 LOCUS 42603 bp DNA linear BCT 12-JAN-2000
 DEFINITION Stigmatella aurantiaca myxothiazol gene cluster, complete sequence.
 ACCESSION AP188287
 VERSION AP188287.1 GI:6635393
 KEYWORDS
 SOURCE
 ORGANISM
 Stigmatella aurantiaca
 Bacteria: Proteobacteria: Deltaproteobacteria: Myxococcales;
 Cyctobacteriineae; Cyctobacteraceae; Stigmatella.
 REFERENCE
 1 (bases 1 to 42603)
 Sliakowski, B., Schaiter, H.U., Ehret, H., Kunze, B., Weinig, S.,
 Nordstedt, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S. and
 Miller, R.
 New lessons for combinatorial biosynthesis from myxobacteria. The
 myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca
 DM4/3-1
 JOURNAL U. Biol. Chem. 274 (52), 37391-37399 (1999)
 MEDLINE 20069734
 PUBMED 10601310
 REFERENCE 2 (bases 1 to 42603)
 Sliakowski, B., Schaiter, H.U., Ehret, H., Kunze, B., Weinig, S.,
 Nordstedt, G., Brandt, P., Blocker, H., Hofle, G., Beyer, S. and
 Mueller, R.
 Direct Submision
 Submitted (20-SEP-1999) NBI/UK, GBF, Mascheroder Weg 1,
 Braunschweig 38124, Germany
 FEATURES
 source
 Location/Qualifiers
 1..42603
 /organism="Stigmatella aurantiaca"
 /mol_type="genomic DNA"
 /strain="DM4/3-1"
 /db_xref="taxon:41"
 1..834
 /gene="mtaA"
 1..834
 /gene="mtaA"
 /note="Pant transferase"
 /codon_start=1
 /transf_table=1
 /product="mtaA"
 /protein_id="AAFI9809.1"
 /db_xref="GI:6635394"
 /translacion="MPTSPALPLKLPDEVHVIVPERITRPGLLSYRALLDPC
 ERDQKQFVFERHLYVLSHVALVRLTSLVAPAPAMGFSANQGRPERIGEXPW
 LRRLSHITDMALCANVARDYGVADVEDTFRSGTVEIASFAPAVASLRALPVSG
 QREFFDVTWKELKAYIVAKMGVSLPUDQAFVSGSLSTRISFDPLVDEPSQWV
 RPRFSORHAAALAVRPSAPLVTRFQRTVPLDDAPAEYLRRRIQPLRMGVGG
 G"
 1388..13399
 /gene="mtaB"
 /note="PKS"
 1388..13399
 /gene="mtaB"
 /note="PKS; MtaB loading domain and module 1: polyketide
 synthase domains: ACP, KS, AT, AT, DH, ER, KR, ACP, MtaB
 module 2: polyketide synthase domains: KS, AT, DH, KR,
 ACP"
 /codon_start=1
 /transf_table=1
 /product="mtaB"
 /protein_id="AAFI9810.1"
 /db_xref="GI:6635395"
 /translacion="MSTREHVGVGRSTGQLRDWLVQVAVSVCGLBAHTVTRTPFS
 LGVGSVPATRLIALELARPLSPVTLVWHPITIDKPARLSCAEPRAEMHAGPEPA
 AVDGAELIALVGLACRPKAPDAKAFWRVQLQEDCVTSVSNRMDSLVYDPAVGP
 GRTVTRGAPLDSVDFDVFEGISPREACQLDPOQRLMELSEWALELDGIAVPRVLR
 GSRTGVGTGMVMDVADLQTIHNAAPLHTCTGAGGI IANRISVALGLESPSLAIDT
 ACSSSLVAIHEACRLSRGSSSLAAGVQVNLSPRFMAVSAFGALAPGRCSTPA
 RASGYVRAGGCVTLKPLSRALADDAITCLIRGSVYVNDGASNGLTANPRQRETL
 LRDAVYRAGMDPAQVHVVELHGTGTLEAGALGAVNGAAREARPLHVGSKTN
 IGHLEAAAGIASIVKVALSTRHQLPASLHFEQNPRIIPDDILRLVQTLAPAPAK
 DPAVAGISAFGVTGNVVEEHRGSRFAAVVSTPRAQSPRHPTAELFAGAGSA

WKMGRELLGAEPVPRAREECDAAPRLPGSWSLVLEALLAPVRSPREGROPLIVL
 AFOIISLAIRMSWGI EPRATVGY SVGENAANA AVAGLITTEEPALATHTITQVKVAG
 GVMGVGLSGVAERLAPYGRIS IISENGPRSTGGEAAALQDFITAGTAAGTIF
 ALVLEIDFYAH I PVEAL KPFLQSCPVLPSPGRVPMVTSVGA FVCGITGATWYMA
 SNREPRPATMRLTLEGGYDVLVDPNPH IHEHS VLENVRA GSAFASIEASMGRS
 PVEGHRALITLOSGAGPVRWERVVDVPGACSSAPVREL PVASARTQALREBADR
 IAEOLIRDAASITDLCTYAGARSHHHRLLVAVETRALREGLAEPAREGIRNPS
 QCMVRGEPAPRLGVPSSGSPFLSSVOL FEEBPSFRKAFRSERLPRACQASLITE
 ALHARESOPHPRTEI EOP LFAFOVALLAWESWIGIRPDVAVGSEVISAHVAGV
 LSEBALSLVRSRLMGRAGQAGMAVEASPARIEPLGCGDGLVIAAENGPRL
 VVSGALALASLRLDELITQGLRVRVLNVAVAFHSQMAPYLQDLPVLELKPPEPL
 PLYTSHGRSLASPGEDTHMRANVERVLPAPEAMADHGAFVESIPRVLLOP
 LSRIIDASPASGASALFSTRBRAPRALIGNLITLTLGHRSWKVGPDLRGVTEL
 PSYMNOROCHEFDAAPOPLAEVPHLGRVNPAPPGQVFEATPRSDRLQILSDRLY
 GIVVAGASMLGMLVAAQOLWGEAPCALQDVLFIQPLVLDTEORLVAS IATGAG
 RLISVAGQGGEGTSSHWI EHA7GRLERGAEETEGAREVLSVEARCRFLASAEFYE
 GIARQGLGIRQFRLVLEALMOGDRERFGLRVERPEDGADYRFHGLDSCFOYGA
 VAGPASIEGLAFVPIGVGRFALRGEARPPRFCHLVLRGSGKANESVDTLLDESG
 LVAAVESGLVIRRAPREVMRLRRDVTAPVLTREMOQOMLSGASOAGPAMVLLDRG
 GLAARLEBOLSGRSRVTLAABENAOSSLGRIEMIDAGLACAGVLDRLGIDRESS
 DAVQAVKSVQOLVTRRAEALRLVLTGTQVTSQDSGLAFBQPLMGDNVIALE
 HPELCQRIDLPSPGAPDAQLAEALAPGREDRVAFREGRILVARTLTPAVRA
 ALPLEGDAYLIVGGLGIGLEIARMMVGRGARHVLNMGRSAPSAHQETLALADAG
 ATISVALGNVAREALAVLGDIRAMPPLRGVMAAGVLDGLVROSEERLAVLA
 PRIOGAMNLRHRETVGRDLDFVLFSVSFVTSQOAGYASANAFLDALHRRAOGL
 TALSTINMGMDVGMMAARLPBPAKSSMSRGLGLAPASGLAERLISSEAOVVA
 GASEBAKCDPAPFFDALSRRKPEASRPVSGALVAGAPRELARLLSLLOVQATV
 LCLPPSAPIDPPRLOELGLDLMALQFNNSLAASIGRLPATLLFDHETLEKLEFL
 AADLTPASEASPAKSRPVKLPARDAGAPPLAVVGMACRLPGSVASPEEFRLAMG
 VNAVEVPKDRMSLDTWYDPDETPGKMSRHGGLFDEVPDPAFPFGISAEASMSD
 POORLLELGMELSGAISAGIADOLKSGSPVFLGICSDYALTEINSPPTNIAVSG
 SGGVTVAGRIAYALGLEGPTLAVDTACSSVLAHILACSLRNGECILAYGSSNL
 LLSPRKSVTFSTKALSPDCAKADPSSANGVYRGAGVLYKRLSEBALADDTLS
 VWRGAIVNODGKSNGLTAPSAQERVIESALRQGVSPLDIVSVYEAHGTSLSDPI
 EVOALAAVNGDRPSAPLIGSVKTNVGHLEAGVSLIKAILALRHELRPSLHF
 KTPSPHIMDELPEVVISHRPMRVGKRRILGVSGFSGTNHVLIEAPENAGPR
 SKAPSIPEBELVLSARAPALOEAVRISAOISGRAGDAAMLHDVYCSANCRSH
 HENRLAVVARSSTEAALAEATLRCGLAHPDNOGYALRGAPRVVLTPLTNGCGMCPRT
 GHALLEBEPFRSALOADEVTLKGLAGMSITELQRETRFPQMAISQPAALDSAL
 LALLRSWGIEPDVAVGSSIGEVAAVAACVLTLEEMRTVARSRLIHAVAKGAMAL
 VLSSEBATEVEAAAOQLVAALINSQNTSLSGEAGMDELGALTAGVSCRIIGL
 DYASHPAVEPYIDELKQALQGLRQPPALIPFSVTYGLAAGDFNSYMANRLOP
 VLAAPVLTILAREDTLFEVLEPNAMVAPVOCFOAVKESATLLSTLRGSEBHTL
 RLNVGLFIAGVTPAMNRLFPBGGSIVSLPSYPMOKERYVIAISPTSARLQLPAPSGV
 LHGRVRSPLASLIPETDLSYKMPRLREHLHGLAVAGAAHVALLASALEIKGTA
 PVRLADIVPRVLKEGVENPVSVVSGEQGKSLRIQSPQSSPEVNTABEVL
 PASSOVEGAQVLAELRLARCPNGLEVSVPYKAMABEGIGLRFRQWEOALQGDGEA
 LGRMAASADQPSAPLHPGLIDSLFVLRAPVPIGRDGVPIPIGIEGLCFIAP
 RGVLANHVAIDLPREGSRDGETLSADPFLFESGKVAMIDGLIVKAPRAAPORHAPD
 LSEKLFQVWQSMPLAPAYERGPMILVADRPAGEBLAARLKRGQOVCHVLAABESA
 VTGELLOVBEREALAEARGVYVLSALAEBOESPQKQALLEGSCAVOYLQAIK
 LERGTFRMLVTRGQALLESFRLALQAPLWGLAVRDEHPELRTCTLIDLPDP
 AADGVCLFQELTTPREDRIAMRGARPAQRISLEAVEPERTAPISDGYVLLTG
 LOGICRCARMLVDAGARHIALIGRSGPEARSVLEALEAGVTVRLOQADVAAREO
 VSKVLSALREVDMPPLRGVLAAGLDDALLVRODRRIQROLOPKVAGALNMDLEERG
 DLDPFLVLESSIAAMGREGGNYSAANAFLALAPAPAOGLPAHSIALMGMAVGM
 VNAABERSAHIALAIGTESIPDPBGVAIILRLILRASAHGITQIDTTRQQAASHGA
 PLLAGLKOQAAAPSDOSSEEAADIQELIRAIPEERQKAEAFLEREVRILIDLPRLQ
 ADGRPLVLRFGFDSIMSLKRLKARLESTIGLISAAKLISGLSIDKLVALVLRITLQ
 PTPVSTDDQMEBFRF"

gene
 CDS
 13537..17409
 /gene="mtac"
 13537..17409

/note="peptide synthetase; peptide synthetase domains:
 heterocyclization, adenylation, PCP note: motif with weak
 homology to putative NAD binding proteins from *Pyrococcus*
 species located behind PCP; this motif may be involved in
 thiazole formation from the proposed thiazoline
 intermediate of the biosynthesis of myxothiazol"
 /codon_start=1
 /transl_table=11
 /product="MtaC"

/protein_id="AAPI9811.1"
 /db_xref="GI:6635396"
 /translation="WIELLRPKXAVPDPAARQPEFPPLTDIOETTFVGRGTPGISGMS
 CHLHEIDTRMDVARSRAWQRIIEOHVMRLSVLPGEORVLESVPFSI PVLDLR
 GOADLPESTKLAELRQOLSTSTSPGQMFTELATLIDBERTDIHPDFDITDASA
 VLTLEERREYLLSDRLPKOVPSPRHVALEAARLTDSPROAEANWNRLLMDMP
 APLPVATEPRRLRSPRRSLRGLDADPTNVLKERANAGLITSGVCAFAEVLGA
 MESPRPTLNTLRRRLIHPEDVRLGDPFTNVLKERANAGLITSGVCAFAEVLGA
 DLERTHSVGMQRQAKMKGNALHPVTSISLGHRSRFGGSPSWLGEVYAL
 TOTQPMMDLHOVREBAGLYCHMWSPEGLPEGLAEFAEYVITDLSLADNLMEL
 GTTVRLPEAQARRETFNATYASIPERLDGLFLAQEOGQRNATIEPALMBEER
 LGOASPLAHLAEGARNEELTAVLEKMGQIVAVTAIHAGSAVLEPIDALPEER
 RLIEEGOVKVLLESHAEBSLSPASRYVAIVADTARAPRLPARASMACTYR
 SSGTRPRGVKIHERRAANTLIDINERGVCPARVLAISGIDSLVYDIFGSLIAG
 AATVPQEVPRDPAHMQVLRQGVTTMWSPTLMEALVDOERAGLLEPTRLVIA
 LSGWMIPTVLDRLRRLSRDRLVSLIGETASLWSIHPIGRVDPARSIPYGRPL
 NQRFVLDKEKLAPPDHPGGLYIGVGLAQGYRDEETARSFENHRTGERLYRTG
 DLGFEPLPEGEIKRQFOVKYIAHRIEELGEIYAVLQRHPSIIEAVALAAGERTSR
 LVAVLVPVEGGEAPSDVLRSEFLAALPOYVPAVPLPETHLPLSNKGVELIKSLML
 VLQSRASGLDRDAGKILSALMRLVSPVLRPSLEPAGILQIGATSVELIKSLML
 EOEGRSPRTTELISLRNASEITAYYERQGAERARSDEALVPAGIRRDACAVTA
 QLEAPDEARQCYFKRVSHRTSAEVEARVLAELSLCSPLKEBQQLVQOSGAG
 GYGVQYTLSTATGRVRLDAGAYYDPLRHTLVOLGASALPALDSANAAADSA
 AFSYLIVDNRVVISQRYGERRDIALIEAGMALLETBRAESDLGICQLEBELRFEGL
 SKALRLEEHVYLGVLGGGLAMEGAL"
 17406..27281
 /gene="mtad"
 17406..27281
 /note="PKS/NRPS; MtaD module 1: peptide synthetase
 domains; heterocyclization, adenylation, PCP note: motif
 with weak homology to putative NAD binding proteins from
Pyrococcus species located between adenylation core
 regions A9 and A10; this motif may be involved in thiazole
 formation from the proposed thiazole intermediate of the
 biosynthesis of myxothiazol; similar to the motif found in
 MtaC; MtaD module 2: polyketide synthase domains: KS, AT,
 DH, KR, ACP"
 /codon_start=1
 /transl_table=11
 /product="MtaD"
 /protein_id="AAPI9812.1"
 /db_xref="GI:6635397"
 /translation="WSTSEFIVYUOQKIELMEIGDTRLPAPRQHLTEPRLTAKDK
 KEQLEHRSASRANLOTVASPRDASRPAPLTDIONVWYVGRGAPAGVAAH
 SYELAFEDTDLPRLESTLQRLIEHMLRVNVPTEGQVILKHYPTFTLTAADLGA
 PASQIEAHFEALRELSHQVLPTRFPLFDRVAVKLAGRVYHNTSPDLTAADFSVO
 ILIBQCAELYNPEAOVLPFRKTRDVCVLAAGADPESRAITNSLAWNRLELPA
 APDLPLVDTDVKQCHQFRRKAGLDQSMRAFPNNAKAGITSMALNARVLELPA
 YSRNRLTNTLTFNRLPLIEDVRITVGDPTFSGILLEDGKSPSFABARLQOQFI
 NDLSHANSVYKVRLEAIRLGRMAGTAMPVPTLSLSETGRSLSVSGKVIIDVISO
 TPQVWLDQVLELDGGIYISDSDALFPRALITAVPESYRLLHRSSDSANDALS
 RQPLPLPAOLARAEVNAATERVSOERIESLFLQAEPRGVAIVTAGQVLSYGLER
 RSARIALAAGKVPDQVAVIAVEKGEVVAALATIRAGAAYVLPSPSLPRLHA
 LLEERQADVLTQSLAESLSLWPEGRRLSVDRDEFLAPRGALPVQGNMLAVIYT

Query Match 22.3%; Score 944.8; DB 1; Length 42603;
 Best Local Similarity 54.1%; Pred. No. 3.5e-133;
 Matches 2269; Conservative 0; Mismatches 1807; Indels 120; Gaps 12;

QY 14 AGCTTCGACAGCTCGACAGCACCGAGTATCAAGCTCGCGCGCCAGTGGAGAGCCCTCC 73
 Db 17419 AGTTCAATGCTATATCTCCAGAGAGGCAATGAACTGTGATCGAAGGAGCAACGCTGC 17478
 QY 74 AGATACAGGCCCCCAAGAACGCCCTGAACCCGAACTCTGCTCGAATCTCCAGACA 133
 Db 17479 GATTCCGGCGGCCCTCGCCACCTCAAGCCGAGCTCGACGCACTGAAGGACAAGA 17538
 QY 134 AAGACAGATCTGACGATGCTCCCTCA-----GAGACTCCCGCGAGAAAT 178
 Db 17539 AGGACAGACTCTTGAGAGATCTGCGTTCCGAGAGCGGCTCCGCGAGATCTTCAAGACCGTGG 17598
 QY 179 CCATGTCGCCGCCCGACCGGAGCGGACAGCTCCGTTCTCTCAAGACATCCAAAGAT 238

Db 17599 CGAGCCCGGATGCGGCGTCAAGCTTCCGCTTCCGCTCAACCAATCCGAATG 17658
Qy 239 CCTATGCGTGGGCGGACAGAGGCTT---TAGGCTCCGAGGGGATCCAGCGCTATC 295
Db 17659 CCTATGCGTGGGAGGAGGAGGCTTCTGATGCGGGCGGCTCCAGCGCAAGTTACC 17718
Qy 296 GCGAATACGACTGATACGATCTCGACGTGCGGAGCTGAGCGCGCTTTGGAAGTCG 355
Db 17719 TCGAGCTGGCGTTGACAGCGCTCGATCTGCGCGCTTGGAGAGCATCTTACAGCGGCTCA 17778
Qy 356 TCGCGCGGACGACATGCTTTGGGCGCCACAGCTGCGGACATGATGAGTATGAGC 415
Db 17779 TCGAGTACACGACATGCTCCGACATGATGCTGCTCCACGCGGACGAGGTGCTCG 17838
Qy 416 CTAAGTGCAGCGCGGACATCGAGATC---ATCGATCTGCGGGGCTCGACGCGAGCACAC 472
Db 17839 AGAAGTCCCGGCTTCACTCCACCGGATGATGATCTGCTGGGCGCGCGGCTCA 17898
Qy 473 GCGAAGCGAGGCTCGTCTGTTGCGAGATGCGATGCGACGCACTTATGACACCGAGC 532
Db 17899 TCGAGGCGCATTTGAGGCGCATCCGCGGGAATCTCCACAGAGTCTTCCAGAGACC 17958
Qy 533 GCCCTCGCTCTATCATCGTCTGCGCTTGGCTGAGACGAGCGGCAACCGCTCTGTC 592
Db 17959 GATTCCCTGCTGATGCTCCGAGGCTCGAAGTTGGCGGTGACGCGGTTCACTCCACA 18018
Qy 593 TCAGATGATCTCATTTAAGTTGACCTAGAGAGCTGCTCATCATTTCAAGAGCTGGC 652
Db 18019 TCAGTGTGATCTCTCAAGCGCGGATGCTGAGGTCCAGATCTGATTCGAACAGTCCG 18078
Qy 653 TCAGCTTACGAGAGATCCGAGACCTCTCTCTGCTGAGAGCTCTGATACCGAGAT 712
Db 18079 CTGCGCTTACCTGAATCCGGAAGCGCACTGTGCTTCAAGCGGACCTTCCGAGATT 18138
Qy 713 A-----TGACTGCGCTGAGATCTCGAAGAACTGAGGCGCATCAAGATGATGG 766
Db 18139 ACTGCTCTGCGCGCCGAGCGCGCTGACCGAGTCCCGGCGCTACAGAACTGCTTG 18198
Qy 767 ATTAATGGAAGGGGCGATGCGCGGCTCCACCTCCGCGGACGCTCGATGAAAGGCG 826
Db 18199 CGTATGCGGAAACCGGCTCGAAGATGCTCCACAGCGCGCGGATGCTGCTGTAACG 18258
Qy 827 ATCCATCTACCTGAGAGAGATCCGCTTCCGACACAGAGCATGCTGCGTGCAGT 886
Db 18259 ACAGCGAGTGAAGGGGACAGACAGATTCGTGCGCGAAGGAGGCTTGAATCAAGCT 18318
Qy 887 CTTGGGCTGATGAAAGCGGCTGTGCGGAGCGCGGCTGACCCGACGCGGCTCATCC 946
Db 18319 CTTGGCGGCAATTCGCAATCATGCAAGCGCGGAGATCAAGGCTCATGAGGCTGT 18378
Qy 947 TGGCTGCTTTTCCAGAGTATCGGCGCTGAGAGCGGACCGCGGTTTACGCTCAACA 1006
Db 18379 GCGCGCTCTATGGGAAACGCTGCGGCTTACAGCGGAGCAACGCGCTTCAACGCTCAAC 18438
Qy 1007 TAAAGCTTTCAACCGGCTCCCGCTCATCCGCGGCTGAAAGATTCACCGGAGACTTCA 1066
Db 18439 TGAACCTTTTCAACCGGCTGCTTGAATCGAGAGTGAAGGAGATCTGCGGAGCTTCA 18498
Qy 1067 CGTGAATGCTCTCTGAGCATCGACACTCGGACAGAAGCTTTCGAACAGCGGCTA 1126
Db 18499 CCTCGGAAATCTGCTCGAATGAGTGGTCAAGACCGGAGTCTTCCCGCGCGGCGC 18558
Qy 1127 AGCGATTCAGAGAGCTGTGGGAAGCGATGATCATCTGCGAGCTAAGCGGTATCGAG 1186
Db 18559 GAGAGCTTCAGGGGAGTTCAATGATCTGAGACACTCGCTGTTTCAGACGGAAG 18618
Qy 1187 TCACAGAGAGCGCGCGGCTCTGGGATTCACAGAGCGCATTTGTTCCCGGCTGTC 1246
Db 18619 TGTCTGGGGAAGCATCTGCGCTTGGCGCTGGGATGCGGAGCGCATGCTTATGCT 18678
Qy 1247 TCACAGCGCGCTTAAACAGCAAGTGTGTGCTCACTGTTGAGAGGCTCGGAACTC 1306
|||||

Db 18679 TCACAGCTTCATCTCGAGACCGGCGCTCATTTAGCGTCTCGAGGG---CGTGAAGA 18735
Qy 1307 CGGTATACACGACGACGACATCTCTGAGCTGCTGATGATCATGACTTACAGGACG 1366
Db 18736 TGTCTGAGCTATCTTCCAGAGCGCTCCAGGTGTGCTCGACACAGGTTTCCAGCTCG 18795
Qy 1367 ATGGGAGCTTCGCTCGCGTGGGACATCTGACGAGAGTGTCCGCGCACTTCTCG 1426
Db 18796 ATGGCGGATCTATTACAGCTGGGATTCGATGATGCGCTTCTCCGCGGCGCTGATCG 18855
Qy 1427 ACGATGCTGGAAGCGTACGCTTTTCTCGGCGGCTTCACTGAGGAACA-----T 1480
Db 18856 ACGCTGTCTTGAAGGCTTGAAGGCTCTTCAACCGTCTTCAAGGATGCTCGGCT 18915
Qy 1481 GGGGTGAACAGTCTGCTGCTTCTCGCTTCCGCTGCGGCTGAGAGCGCGGACGCA 1540
Db 18916 GGGATGCACTTCAGAGGACCGCTCTTCCGCGGCTTGGCAAGCGCTGCCAGTACA 18975
Qy 1541 ACGGACCAACGCGCTCTGAGGAGCATACGCTGCAAGCGCTGTTGCGGCGGCTCG 1600
Db 18976 ACGCACCAAGCGTCCGCTGCGAGAGCGGATCGAATGCTCTTCTGTCAGGCGG 19035
Qy 1601 AGCAGCTGCCATCATGACTGCGCTGCTGTGTGGCGCGGACAGCCTCATCGTGAAGC 1660
Db 19036 AGCTCAACCGGCGCGCTGCTTATCGCGGAGAGCAGATTTCTCTATGCAAGC 19095
Qy 1661 TTTGCGCGCTTGGCGGACCTTGGCGCGGCTGCGGAGCAGAGGAGGACCGCGCA 1720
Db 19096 TCGAGCGCGGCTCGCTGATGATCGAGGTGCTCATGCGCAAGGTGTCAGAGCTGACC 19155
Qy 1721 CATTTGCGCGGTGATGATGAGAAAGCTGGAGAGGTTGTGCGGCTTCTCGCGTGC 1780
Db 19156 AGCTGTGGGATGCTTGGCGGAGAAAGGAGGAGGAGGAGTGTGCGCGCTGCGCATCC 19215
Qy 1781 TCGAGTCAAGCGCGGCTTACGCTGCGATTCAGTCCGACCTTACGCGGAGGCTATCCACT 1840
Db 19216 TTTGCGGTGAGCGGCTTACGCTGCTTCTCATCTGCGCGCGCGGAGGCTTCCAGC 19275
Qy 1841 ACCTCTCGATCATGATGATGAGTAAAGCTCGCTGACGACGACCATGCTGATGGAAC 1900
Db 19276 CTTGCTGGAAGAGGCGGACCGCACTGTGCTTCAACCGACTCGCGCTGAGATGCTGCG 19335
Qy 1901 TGTATGCGCGCGCGGATTCAGCGGCTGCTGTAAGCGAGCGCGCTCGAAGCGAGC 1960
Db 19336 TGGATGCGCGGAGGCGCTGCTGCGCTTTCGCT-----GGATGATGACAGATTTCG 19389
Qy 1961 GCGACAGCTCCGATGATGCTTCAAGACACTTGGAGTCTGCGTATGCTTCA 2020
Db 19390 AAGCCCGAAGGGGCGCTGCGCGGTTCAAGGCAACATCTGCTTCACTCATCA 19449
Qy 2021 CTTGCGGATTCAGAGGTTGCGCAAGGGGATGATGATCATCGGGGTGCGTCAACA 2080
Db 19450 CTTGCGGCTCGAGCGGCGCTTCAAGGCGGATGATGACACCGGGAGCGGTAACA 19509
Qy 2081 CCATCTGACATCAAGAGCGCTTTCGAATAGGCGCGGAGAGAGGCTGCTGCGCTCT 2140
Db 19510 CCTCTCGACATGAAGAGGCTTCAATGTGGGCGCGGAGCGGCTGCTGCGCTT 19569
Qy 2141 CTTGCTGAGCTTGAATCTCTGCTTATGATGTGTTGCGGATCTTGGCGGCGGCTA 2200
Db 19570 CGTCACTGAGCTTGAATGTGCTTCAAGCACTTTCGAGGTTCTTGGCGGCGGCTG 19629
Qy 2201 CGATCGGTGCGGAGCGGCTTCAAGGCTGCGGATTCGCGGCAATTTGGGCAAGTTGATCG 2260
Db 19630 CGATGATGATGCGGAGCGCGGACGCTGCGGAGTCCCGAGCGGAGTGTGCTGCG 19689
Qy 2261 AACGAGAAAGTGAAGCTGTGAACTCGGTGCGGCGCTGATGATGCTGCTGAGC 2320
Db 19690 AGAAGAGGGGGTGAACATCTGGAATCGGTCCCGGCTTGAATGACATGCTGCTGAGT 19749
Qy 2321 ATTTGAGGTGCGCGCGGATTCGCTCGTATGATCTTGGCGCTTTCGCTGCTGAGCGG 2380
Db 19750 TCTCGAGGGAAGCGGTCTGCGCTTGGCGGACTGCTGCGCTGCTCATGAGCGGTG 19809
|||||

QY 2381 ACTGATCCCGGTGGGCTTCCCTGAGAGCTCCAGGCACTACAGCCCGGCTGTCTGTGA 2440
 Db 19810 ACTGATTCCTGGTCAAGCTGCCCCGGAGATCCGTCCTGTCCAAAGACCGGAGCTGG 19869
 QY 2441 TCAGCTCGGGGGGGGACCCGAAGCGTCGATCTGTGTCATCGGGTACCCCGTGAAGAAC 2500
 Db 19870 TCAGCTCGGGGGGGGACCCGAAGCGTCGATCTGTGTCATCTGTGTCATCTGTGTCAT 19929
 QY 2501 TCAGCTTCGTGGGAGAGCATCCCTCAAGCGGCTGTGTCGTCGTCGTCGTCGTCGTCGTC 2560
 Db 19930 TGGGGGGCGGCTGGCGGACATTCGCGTACGCTGCGCCCATGTGATTCAGCCCTTCATG 19989
 QY 2561 TCGTCGATGAGAGCGCTTCGAAACCGCGCCCGCTGTGGTTCGGGGGCACTCTACATTGGG 2620
 Db 19990 TCGTCGATGAGAGCGCTTCGAAACCGCGCCCGCTGTGGTTCGGGGGCACTCTACATTGGG 20049
 QY 2621 GGGTCGGGCTGGCACTGGCTACTGGCGGATGAGAGAACCGCGCAAGAGCTTCCTG 2680
 Db 20050 GCAATGGCTCTTCGCTCGCTGCTATTCAGGAGATCCGCTTCGTCACCGCGAGCGCTTCATG 20109
 QY 2681 TGCACCCCGGAGACCGGGAGCGCTTCACAGACCGGCGATCTGGGCGCTACCTGCGCG 2740
 Db 20110 TGCATCCCAAGACCGGGAGCGCTGTACAGACCGGCTGACCTGGGCGGCTTCATGCGCG 20169
 QY 2741 ATGGAACATTCGAGTTCATGGGGCGGTGAGAGCAACCAATTCAGCTTCGGGATACCGG 2800
 Db 20170 AGGGGAACATTCGAGTTCCTCGAGCGGAGGACTTCAGGTCAGAGTTCAGGCTTCGGA 20229
 QY 2801 TTGAGCTCGGGGAAATCGAGGAAACGCTCAGTTCGATCCGACGTCGCGAGCGGTA 2860
 Db 20230 TCGAGCTCGGGGAGATCGAGGCGCGCTGACTTCGATCCAGCGGCTCTCTCGGTTG 20289
 QY 2861 TTGTGCGCTCGGAAAGACGCGGCGAACAAGCTCTTCTAGCTTATGTGTCCCGAG 2920
 Db 20290 TCAATGCTTGGAAAGACCGGGGGGAGCTCGGCGCTCGTGGCTATGTGTCTCGGAG 20349
 QY 2921 GCAACCGAGAGCGCGTGCAGAGCAAGAGCGGCTCAAGACCGAGC---GAGTCAGAG 2977
 Db 20350 AATTGACGTGAAATGCTCCCGGCAAGCGCGCTCGGCGAGCAAGCTCTCTTGGAGT 20409
 QY 2978 CGAGAGCAACACCGCGGAGAGCGAGCGCTTGAAGCGAGCGAGAGG----- 3025
 Db 20410 CGGCTCCGAGAGCTTGCAGAGTACAGCGCTCTGTGAGAGCCAGGGGCTGGGCTGATCA 20469
 QY 3026 -----TCAATTCAAGCTCTGCTGACACGAGCTCGGAGGAGACCTTG 3067
 Db 20470 CCGATACGCTGGAACGCTGAGATTCAAGCTGAGACACAACTGCGCGGATGACAG 20529
 QY 3068 AGGGAAGCGCGCTCGCTGATCTGACCGGCGAGATCCGGGAGGCGGGGCTGAGCTCT 3127
 Db 20530 GAGCGCGGCTCAAGTTGCGCTCAAGAACCGGAGTTGATGAGGCTTCCCGAAGAGT 20589
 QY 3128 AGCGCGTGCCTGAGCGCTCGCAACGTTCTTGAAGCGCGGATTCGTTGTGTTGAGTTG 3187
 Db 20590 ACATGAGGCGGAGAGCGCGCGAGCTCTCTCGAGCGCGCTCTCTGAGAGCGGCTCA 20649
 QY 3188 GTTCGATTCCTGAGCTGTGAGCAGCGTGAAGCGCGGAGCGCGGACCTTCCCAATTCC 3247
 Db 20650 GCGAGTTGCTGAGCTGTGAGTCAAGATCGCTCTCGAAGATCTCAATGATGCCCAAGTAC 20709
 QY 3248 GTTATTCATTCGCGGCGGACAGTACCGGCTGCAACCTAAGCTATGTCGAATCCGGCG 3307
 Db 20710 GGTATGATTCGCGGCTGTGCTTATTCGAGTGTGATCTGACGTCGAAGCCAGGGG 20769
 QY 3308 GCATCGAGG---CGTGAAGAGGCGCTTATATTAATCAACCGCTTCAGACCGCTTTCG 3364
 Db 20770 GTGTGGGGGGGCTCGGGGGGGGAGCTACTATTAATCAACCGAGCTCAAGAGCTCTG 20829
 QY 3365 TGAAGCTCTCCATCAAGGATTCAGCGCGAGCGAGCGAGCTTCGGCAAACTTCGAGCTGT 3424
 Db 20830 TCGTACCGCGGATTCGCGCATGAGACAGAGCCAAACAGCGCTCCGGAACCGGCGCTCT 20889

QY 3425 TCGATGAAGCGGCGTTCAACCTCCCTGTTGTTGGGAGGATCGAACCCCATGAGTCTGT 3484
 Db 20890 TCGATGCCGCGCGCTTCCCTGTTTCTCTGTTTGAAGACTAGTCCATCCGCGCTGT 20949
 QY 3485 ATGATTCCTGTGCGAGAAATTTTTCCTGTGAGAGCGAGTATATATGCGAGCTCTGA 3544
 Db 20950 ACGGAGATGAGCGGGGATTTTCTGATGTCGAGAGCGGGCTACATGCTCAGCTCTCA 21009
 QY 3545 TGAAGACGCGCTTCTCTGCAACATGCGCTGTCTCGGTGGGCAATTCATTTTGAAC 3604
 Db 21010 TGAAGCTTCCCGCCCAAGATGGGTTGTGCTCCATCGCGGTGATGACCTTCAGC 21069
 QY 3605 AGGTTGCGGCTTTCGACCGCGACATTCGAGCTTACGTCGACGAGATGCTGGGCG 3664
 Db 21070 CCTGCGCTCGAGCTCGCGTTGAGACGACACAGTCTTCTGACAGCTTCTTGAGG 21129
 QY 3665 GCGGCTGAGACCGCGGAGTTCAGCTGTGTACGCTCGTACAGATTCCTACCCAGAGC 3724
 Db 21130 GCGGCGAG-----TGAAGCGGCG 21148
 QY 3725 GCGGCAAGACCGGCGGCGCTCCCTCCGCGCGAGACGACATTCGCGATTCCTGCG 3784
 Db 21149 GCGTCCGCGCAAGCGAGCAGACCCAGCGCGGATGCTTCACTGC-CCGACAGTTGAGAGC 21207
 QY 3785 ACTTGTGAGGACCAACCTACCCGATCATGTCGCTTACAGTCTTCGTGAGCTCGATG 3844
 Db 21208 GTTACTGTGCTGAGAGCTGCGGAGTACATGTTCCGCTCTGTTCTGTTGAGATT 21267
 QY 3845 GCTTCCGCTGACGTCACAGCGGAGGTCGATTCGTAAGGCTTCGCGAGCGGAGATA 3904
 Db 21268 CCTTCCGCTGACGTCACAGCGGAGGTCGATTCGTAAGGCTTCGCTCGGCGGAGC 21327
 QY 3905 CCTGCTCCGCGGATTCGCGGAGCAGCGGCGGAGCGGAGGCTTCGAGAGATTCCTG 3964
 Db 21328 TGTGTGAAGTCAAGCGCGCGCTGCGCGGAGGCTTCGAGAGGCTTCGAGG 21387
 QY 3965 TCGCGGCTGTAAGGAGGTCGCGGCTGAGAGGTCGCGGCTCCAGAGAGCTTCGTCG 4024
 Db 21388 CCGCATGCTTCAGAGGTCCTCAACTGCAAGAGTGAATGTCCACCGCACTTCCTG 21447
 QY 4025 ATCTTGTGCAATCATCATCTGTTGCAATGAGAGGCTTTCGAGAGAGGCTG 4084
 Db 21448 ATCTGCGGCGGAGCGCTGCTTCAATCGTCAAGATCGAGATGAGAGGCGGCTCA 21507
 QY 4085 ATAGGAGATGCGCATCAACCGAGTTCGATGACCGGACCTCGGCTCGCTGCG 4140
 Db 21508 AGCGCATATTCGAGTTCGCGGAGATGTCGCTTACACCAACCGTACGTCGCTGCG 21563

RESULT 11.
 AF210249
 LOCUS 77457 bp DNA linear BCT 29-AUG-2000
 DEFINITION Streptomyces verticillius bleomycin biosynthetic gene cluster,
 complete cds.
 ACCESSION AF210249 AF149091
 VERSION AF210249.1 GI:9937210
 KEYWORDS
 SOURCE Streptomyces verticillius
 ORGANISM Streptomyces verticillius
 Bacteria; Actinobacteria; Actinomycetales;
 Streptomycinae; Streptomycetaceae; Streptomyces.
 REFERENCE 1 (bases 74421 to 74693)
 Du, L., and Shen, B.
 Identification and characterization of a type II peptidyl carrier
 protein from the bleomycin producer Streptomyces verticillius ATCC
 15003
 JOURNAL Chem. Biol. 6 (8), 507-517 (1999)
 MEDLINE 99352421
 PUBMED 10421758
 2 (bases 1 to 77457)
 REFERENCES
 Du, L., Chen, M., Sanchez, C. and Shen, B.
 An oxidation domain in the BlnII non-ribosomal peptide synthetase
 probably catalyzing thiazole formation in the biosynthesis of the

CDS

/codon_start=1

Query Match 17.2%; Score 728; DB 1; Length 77457;
Best Local Similarity 55.9%; Pred. No. 3.8e-105;
Matches 1545; Conservative 0; Mismatches 1185; Indels 36; Gaps 7;

182 TCGTCCCGCCCGCCGAGCGGCGGCGCTCCGTTCTCTTCTCAACAGATCCAGATCTT 241
Db TCGTCCCGCCCGCCGAGCGGCGGCGCTCCGTTCTCTTCTCAACAGATCCAGATCTT 61102
242 ACTGCGTGGGCGGAGAGGCGTTTACGGTCCCGCCGAGCGG---ATCCAGCGCTATCCG 298
Db ACTGCGTGGGCGGAGAGGCGGTTTACGGTCCCGCCGAGCGGCTCTTCCAGCGCTTACCTG 61102
61103 ACTGCGTGGGCGGAGAGGCGGTTTACGGTCCCGCCGAGCGGCTCTTCCAGCGCTTACCTG 61102
299 AATGCGACTGTACGATCTCGACGTCGCGAGCGCTGAGCGCGCTTTGGAAAAGTCGTCG 358
Db AATGCGACTGTACGATCTCGACGTCGCGAGCGCTGAGCGCGCTTTGGAAAAGTCGTCG 61102
61163 AGATCGAGGCGCCGCGGATCGACGTCGACGCTTTTACCGGCGCGCTGCGCGGCGTATCG 61222
359 CGCGCGACGACATGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 418
Db CGCGCGACGACATGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61282
61223 CGCGCGACGACATGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61282
419 AAGTCGAGCG---CGACATCGAGATCATCTGCTGCGGCGGCTGCAACGCGGAGACGCGG 475
Db AAGTCGAGCG---CGACATCGAGATCATCTGCTGCGGCGGCTGCAACGCGGAGACGCGG 61342
61283 AGTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61342
476 AAGCGAGCGTGTGTCGAGATGCGATGTGCGACCGCATCTATGACACGAGCGCGC 535
Db AAGCGAGCGTGTGTCGAGATGCGATGTGCGACCGCATCTATGACACGAGCGCGC 61402
61343 AAGCGAGCGTGTGTCGAGATGCGATGTGCGACCGCATCTATGACACGAGCGCGC 61402
536 CTCCGCTCTATCAGCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 595
Db CTCCGCTCTATCAGCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61402
61403 GCGCGCGTGT---CGACGTCGCGGCTTCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61459
596 GTATCGATCTCATTAAGCTTGAAGCGAGCGGCTGTCATCTTCAAGAGCTGAGCTCA 655
Db GTATCGATCTCATTAAGCTTGAAGCGAGCGGCTGTCATCTTCAAGAGCTGAGCTCA 61519
61460 GGGTGAAGCGGCTGATCTGCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61519
656 GCTTCTACAGAGATCCGAGACCTCTCTCTGCTCTGAGAGCTTCTGTAACGCGGATTAAG 715
Db GCTTCTACAGAGATCCGAGACCTCTCTCTGCTCTGAGAGCTTCTGTAACGCGGATTAAG 61579
61520 CCGGTTACGCGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61579

QY 716 TACTCGCGCTGAGAGCTGTCGGAAGAAGTCTGAGGCGCATCAAGATGATGATTAATCGA 775
Db TACTCGCGCTGAGAGCTGTCGGAAGAAGTCTGAGGCGCATCAAGATGATGATTAATCGA 61580
776 AGCGCGCATGCGCGGAGCTTCCGACCTTCCGCGGAGCGCTTCCGATGAAGCGGATCATCTA 835
QY AGCGCGCATGCGCGGAGCTTCCGACCTTCCGCGGAGCGCTTCCGATGAAGCGGATCATCTA 61640
Db AGCGCGCATGCGCGGAGCTTCCGACCTTCCGCGGAGCGCTTCCGATGAAGCGGATCATCTA 61699
836 CCTTGAAGAGATCCGCTTCCGCGGACACGAGACATGCTGCTCCGCTCGGATCTTCCGAGGTC 895
QY CCTTGAAGAGATCCGCTTCCGCGGACACGAGACATGCTGCTCCGCTCGGATCTTCCGAGGTC 61700
Db CCTTGAAGAGATCCGCTTCCGCGGACACGAGACATGCTGCTCCGCTCGGATCTTCCGAGGTC 61759
896 GATTGAAGCGGCGGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 955
QY GATTGAAGCGGCGGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61760
Db GATTGAAGCGGCGGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61819
956 TTTCCGAGGATATCGGCGGCTGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1015
QY TTTCCGAGGATATCGGCGGCTGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61820
Db TTTCCGAGGATATCGGCGGCTGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61879
1016 TCAACCGCTCCCGCTCCATCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1075
QY TCAACCGCTCCCGCTCCATCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61880
Db TCAACCGCTCCCGCTCCATCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61939
1076 TCTTCTGAGATCGACACCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1135
QY TCTTCTGAGATCGACACCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61940
Db TCTTCTGAGATCGACACCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61999
61940 GCTTCTGAGATCGACACCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 61999
1136 AAGAGCACTGTGGGAGGCGGAGTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1187
QY AAGAGCACTGTGGGAGGCGGAGTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62000
Db AAGAGCACTGTGGGAGGCGGAGTCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62059
1188 -----CGAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1240
QY -----CGAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62060
Db -----CGAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62119
1241 TGGTCTGACGAGCGGCGGCTTAACGAGAGTGTGTCGACCTGTTGCGAGAGGCTCG 1300
QY TGGTCTGACGAGCGGCGGCTTAACGAGAGTGTGTCGACCTGTTGCGAGAGGCTCG 62120
Db TGGTCTGACGAGCGGCGGCTTAACGAGAGTGTGTCGACCTGTTGCGAGAGGCTCG 62179
1301 GAACTCCGAGTGAACACGAGCGGAGCTCTGAGCTGCTGAGATCATGAGCTTAA 1360
QY GAACTCCGAGTGAACACGAGCGGAGCTCTGAGCTGCTGAGATCATGAGCTTAA 62180
Db GAACTCCGAGTGAACACGAGCGGAGCTCTGAGCTGCTGAGATCATGAGCTTAA 62239
62180 GAGAGCGGCTGTAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62239
1361 AGCAGATGAGGAGCTGCTGCTGCGGCGGAGCATGCTGAGAGAGTGTCCCGCGGAGC 1420
QY AGCAGATGAGGAGCTGCTGCTGCGGCGGAGCATGCTGAGAGAGTGTCCCGCGGAGC 62240
Db AGCAGATGAGGAGCTGCTGCTGCGGCGGAGCATGCTGAGAGAGTGTCCCGCGGAGC 62299
62240 AAGAGCGAGGAGTGTGCTTCAACTGAGAGCGGCGGAGAGCTGTGCTCCCGCGGAGC 62299
1421 TTTGAGAGCATGCTGAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1474
QY TTTGAGAGCATGCTGAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62300
Db TTTGAGAGCATGCTGAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62359
62300 CCTTGAAGAGATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62359
1475 AAGCAATGAGGAGTGAACAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1534
QY AAGCAATGAGGAGTGAACAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62360
Db AAGCAATGAGGAGTGAACAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62419
62360 AAGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62419
1535 GCGCAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1594
QY GCGCAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62420
Db GCGCAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62479
62420 GCAACGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62479
1595 GGGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1654
QY GGGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62480
Db GGGTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62539
62480 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62539
1715 GGAACATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1774
QY GGAACATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62600
Db GGAACATGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62659
62600 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62659
1775 CCGTCTGAGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1834
QY CCGTCTGAGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 62659

Db	62660	GCCTCTGAGATCGGGGGCGGGTCACTGCCCCCTGAGACCCCGAACTGCGCCGCAACGCGC	62719
Qy	1835	TCACATCTCTCTGATCATGATGAGGTAAAGCTCGTGTGACCGCACGATGCTGTGATG	1894
Db	62720	TGCTTCACCTCTGTAACGGCGCGCCGAAGCCGCCCTCTCTCTCAACGAAACGCCCTCTGCG	62779
Qy	1895	GCAATCTGTCAATGCGCCCGGGGATTCACCGGCTGCTGTGAGCGAGCGCGCTCGAAG	1954
Db	62780	ACAGCGCTCGCGCTCCCGTCCGGCGTCAAC---CGTGTCTCGCGTGGAGACGACGCGCGCC	62836
Qy	1955	GCGACGGGGAACGAGCTCCGATGATGATGCCATTCAGACACTTGGGATCTCGCGATATCA	2014
Db	62837	TGAGACCGGACGCGCGGCCCGCTGAGAGCGTGCAGAACTTCACCGACCTGCGCTGACCA	62896
Qy	2015	TCTACACTCCGGATTCACAGGCTTGGCCCAAGGGGTGATGATGATCATTCGAGGATCGCG	2074
Db	62897	TCTTACCTCGGGCTCCACCGCGCAACCCAGGCGCTCATATGACACACTCGCGCGCG	62956
Qy	2075	TCAAACCAATCTTGAGCATCAACGAGCGCTTCGAATAGCGCCCGGAGACAGGCTCTGG	2134
Db	62957	CCAACACCTTGAAATGCGTCAACGCGGCTTCGCGACCGCGCCCGGAGACGCGGCTCTCG	63016
Qy	2135	CGCTCTCTCTGCTGATGATCTCTCTGCTTATGATGTGTGGGATCTCTGCGCGCG	2194
Db	63017	CCGCTCTCTCCCGAGCTTCACCTCGCGCTTACGACTGTGTTGGGCTGTGCGCCCGC	63076
Qy	2195	GCGGTACGATCGTGGTGCAGGACCGCTTCAAGCTGCGCGGATCGGCGCATTTGGGCAAGT	2254
Db	63077	GCGGACCGTGTGTGTCTCCCGCCACACGACCGCGCGCGGACCCCGACACTGGCGGACG	63136
Qy	2255	TGATTCGAACGAGAAAGTGAACGCTGTGGAATCTCGTGCAGCGCTGATGCGATGCTCG	2314
Db	63137	TGATTCGCGCGGACGCGGATCACTCTGTGGAATCTCCGCTCCCGCGCTGGGACCTCTGCTCA	63196
Qy	2315	TTCGAGCATTTGAGAGGTGCGCCCGATTCGCTGCTAGTCTCTGCGCTTTCGCTCTGA	2374
Db	63197	CCGAGTACGCGGAGGCGCTCCCGCCCGACGCGCTTCGCGACCTTCGCGGCGGTGCTCTCTCA	63256
Qy	2375	GCGGCGACTGATCCCGGTGGGCGCTGCGCTGCGAGCTCCAGGCGCAT---CAGGCGCGGCG	2431
Db	63257	GCGGCGACTGATTCCTCTCGGATTCGCCCACTGCCGACCGGATTCGCGCGCTTCGCCGCCGCG	63316
Qy	2432	TGTCGTGATCAGCTTGCGCGCGGCGCAACGAGCGTGCATCTGCTCCATCGGATACCCG	2491
Db	63317	CCACCGTATAGCGCTCGCGCGCGCGGACGAGCGCTCATCTGCTCGGTCTGTGACAGA	63376
Qy	2492	TGAGGAACGTGCATCTATCTGTGGCGAGCATCCCTTACGCGCGCTGCGCGCAACAGA	2551
Db	63377	TTCGGAAAGTTCACGAGGCGTGCAGAGCATCCCTTACGCGCACCCCATGCGCAACAGC	63436
Qy	2552	CGTTCACGCTGCTGATGAGGCGCTGCACACCGCGCCCGGTCTGGGTTCCGCGGCAACTCT	2611
Db	63437	GCGTGAAGTCTTGCAGAGAGAGTGTGCGGCCCGCGACCTGAGTGGGTGCTCCGCGAGCTGT	63496
Qy	2612	ACATTTGCGGAGTTCGCGGCTGCGCACTGGGCTACTGGCGCATGAGAGAGAAGCGCGCAGA	2671
Db	63497	ACATTCGGGGGACCGGCGCTCGCAAGGGCTTCTGGCGGACCCGGAACAGACTCTCTCG	63556
Qy	2672	GCTTCTCTGTCACCTCCGAGACCGGGGAGCGCTTACCAAGACCGGCGATTCGGGCGCT	2731
Db	63557	GCTTCTCCGTCACCCCGGAGCGGGGCAACGCTGTACCGACCGGGGACCTTGCCTCGCC	63616
Qy	2732	ACCTGCGCGATGAGAAACATTCAGTTCATGGGGCTGAGAGACACCAATTCAGCTTTCGCG	2791
Db	63617	ACCTTCCCGAGGAGCGCTGGAATTTCTCGGGCCGACGAGACGACCAAGTGAAGTTCGCG	63676
Qy	2792	GATACCGGCTGAGCTCGGGGAAATTCAGAGAAACGCTCAAGTGCATCCGACGTACGCG	2851
Db	63677	GATTCGCGGTGAACTGGGCGAGGTTCAGAGCGGCGCTTCGCGGACCTGCGCGAGCTCGCG	63736
Qy	2852	ACGCGGTGATTTGCGCGCTCGGGAACGACGCGGCGAACAGCTCTTTCAGCCTATGTG	2911

Db	63737	CCGCGCGGTGATCGCACCGGATGACCCCGGGGCGACCGCGCGCTCTGTCGCGCTTCGCG	63796
Qy	2912	TCGCGG 2917	
Db	63797	TACCGG 63802	
RESULT 12			
AL646066/c			
LOCUS			
DEFINITION	AL646066	189050 bp	DNA linear BCT 11-APR-2003
ACCESSION	Ralstonia solanacearum		
VERSION	segment 10/79		
KEYWORDS	AL646066		
SOURCE	AL646066.1	GI:17428714	
ORGANISM	Ralstonia solanacearum		
REFERENCE			
AUTHORS	1		
	Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,		
	Arlat,M., Billault,A., Brotier,P., Camus,J.C., Cartolico,L.,		
	Chandler,M., Choiane,N., Claudel-Renard,C., Cunac,S., Demange,N.,		
	Caspin,C., Lavie,M., Moisan,A., Robert,C., Sautin,W., Schlex,T.,		
	Signier,P., Tineault,P., Whalen,M., Wincker,P., Levy,M.,		
	Weissenbach,J. and Boucher,C.A.		
TITLE	Genome sequence of the plant pathogen Ralstonia solanacearum		
JOURNAL	Nature 415 (6871), 497-502 (2002)		
MEDLINE	216181879		
PUBMED	11823852		
REFERENCE	2		
AUTHORS	2 (bases 1 to 189050)		
TITLE	Direct Submision		
JOURNAL	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston		
	Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie		
	Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,		
	BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean		
	Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS		
	118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA		
	URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,		
	Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,		
	F31326 Castanet-Tolosan Cedex		
	INRA, BP27, F31326 Castanet-Tolosan Cedex		
	Christian.Boucher@toulouse.inra.fr		
	http://sequence.toulouse.inra.fr/R.solanacearum.html.		
COMMENT			
FEATURES			
source	1		
	1..189050		
	/organism="Ralstonia solanacearum"		
	/mol_type="genomic DNA"		
	/strain="GM11000"		
	/db_xref="taxon:305"		
	170..3325		
	/gene="RSC1698"		
	/note="synonym: R502885"		
	170..3325		
CDS			
	170..3325		
	/gene="RSC1698"		
	/note="synonym: R502885"		
	170..3325		
	/function="elements of external origin, phage-related		
	functions and prophages"		
	/note="Product confidence : probable		
	Gene name confidence : hypothetical		
	predicted by Codon usage		
	predicted by Homology		
	predicted by Framed"		
	/codon_start=1		
	/evidence=not_experimental		
	/transl_table=11		
	/product="PROBABLE BACTERIOPHAGE PROTEIN"		
	/protein_id="CAD15400.1"		
	/db_xref="GI:17428715"		
	/db_xref="SPTREMBL:O8XYR0"		
	/translation="MRN11GYGGKGGSSSSSTPTAPDLSVAVARVLVDSBEIA		
	GLNGLQSYLEGTPLANADTLNPNVAADVPPSQDEALPGFPPSVESBAVVEL		
	PLTWMPVRAINTOLSAVRVQLSVPALSRADSSGNGINGRYREVAIDILAVGGALQOV		

Dh 137929 TCGACGAGCTTACTGATTCGACCGCCAGGGCCACTACGCGCTGGGCGCGCTTCC 137870
Qy 287 AGCCTATCGGAAATACGACTGTACGAGATCTGACGTGCCAGGCTGAGCGCGCTTTC 346
Dh 137869 ACGCTTACTGAGAGCCATGCCGCGACGCGCTGACCTGGCTGCTTCAGACCTCTCT 137810
Qy 347 GGAAG---TCGTCCGCGGCGACGACATGCTTCCGGCCCAACGCTGCCAGCATGATGC 403
Dh 137809 GGAATGCCGTGTGCGCGCCACGCGCGCTGCGCTGTGTATTCAGACGAGCGGCGAGC 137750
Qy 404 AGGTATGAGCTTAAATTCAGACCGCAATCGAATCATG--ATTCGCGCGGCTCG 460
Dh 137749 AGCGATCTCGCGGGGTGCGGCTTACCGATCCGGTGGCAAACTGGGCGCGGCA 137690
Qy 461 ACCGAGCACAAGGGAAGCGAGGCTGTGTGTTGAGATGCAATGTCGACCGCATCT 520
Dh 137669 CCGCGAGGCTTCGCGCGCGCGACGTGCGACGACTGGCGCCAGGCAATGTGCGACGAGTGC 137630
Qy 521 ATGACACCGAGCGGCTCGCTCTATTCAGCTGTGCGCTTCCGCTGAGACGAGCGCAAA 580
Dh 137629 TCGATGCGCGCAATGCGCGCTGTTCAGATGCGCGCACGCACTGCGCGGCGGCGCA 137570
Qy 581 CCGCTCTGTGCTAGTATTCGATCTCATTTAAGTTGACCTTACGAGCGCTGTGCATCT 640
Dh 137569 CGCGCTGACATCGGATTCGACATGTATCAACGACGCGCACGCGCGGCGGATCATCT 137510
Qy 641 TCAAGA-----CTGGCTCAGCTTCTACGAAAGATCCCGAGACTCTCTCCCTG 688
Dh 137509 GGGAGGACTGCGCGCGCTGTACCGCGCGCGGCGGACCTCGAACGCGCGCGCTGCGC 137450
Qy 689 TCTGAGACTCTGATCCGCGCATTTATTTACTCGCGCTGAGTCTGCAAGAGT----- 743
Dh 137449 CATTCGAGATCAGCTTCCGCGCATACGTGCTGGGAAATTAAGTGCACGCGAGCGCGC 137390
Qy 744 -TGAGCGCATCAACGATGATGATTACTGAGAGCGCGCATGCGCGAGCTCCCATCTC 802
Dh 137389 GCGCGCGCGGAAATCGGCAAGGCTTACTGCTGCGGACGTCCAGACCTGCGCCCGC 137330
Qy 803 CGCGAGGCTTCCGATGAGGCGCATTCGATCCCTGAAGAGATCCGCTCCGCGACA 862
Dh 137329 CGCGCAATGCGCTGCGCGCGCGCGCGCGCTGCGACCGCGCGCGCGCGGATTTCTGCGGC 137270
Qy 863 CGGAGCATGTGCTGCGCTGCGACTCTGAGGTTCGATTGAGCGCGGCTGCGGAGCGCG 922
Dh 137269 GCGAGCATGTGCTGAGCGCGCGCGCGCATGTGCAAAAGCTGCGGACCGCGCGCGGATGG 137210
Qy 923 GAGTGAACCCGAGCGGCGTCACTCTGCTGATTTTCCGAGGTGATCGGCGCTGAGCG 982
Dh 137209 GCTGCAACCGCGGCTTCTGCTGATCGCGGTGTTGCCGAAGTCTGTGCGCTGAGACA 137150
Qy 983 CGAGCCCCCGGTTTAACTGCAACATTAAGCTTTCAACCGGCTCCCGCTCATTCGCGCG 1042
Dh 137149 CCGAGCGCGGTTCACGCTCAACCTCACACCTTGCACGCTTGCCTGCGGACCGCGACG 137090
Qy 1043 TGAAGATATACCGGCGGACTTCAAGTGTCTCTGGAACATTCGACCACTCGCG 1102
Dh 137089 TGGCGCGCTGTGGCGCATTTACCGGCTGACGCTGCTGCGTGAATGCGCGCGC 137030
Qy 1103 ACAAGAGTTTGAAACGCGGCTTAAGCGTATTCAGAGACAGCTGTGGAGACGATGATC 1162
Dh 137029 CCGTCCCTTGGCGCGAGCGCGCGCGCTCAACGCGGCGGTGTGTCAGACCTGCGAGC 136970
Qy 1163 ACTGCGAGTAAAGCGGTATTCAGAGTCAAGAGCGCGCGCGCGGTCTGTGGGATTCACAC 1222
Dh 136969 ACCGCGCTTACGCGCGCGGTGACGTGTGCGGAAATGGAACCGCGCGCGACGCGCGG 136910
Qy 1223 GAGGCGATTGTTCCTCGGTGTCTCAGAGCGCG-----TTACAGCAAGTGTG 1276
Dh 136909 ATGGGATTCATGCGCGGTGTGTTTACACGCGAGCTGGCATGAGCAACCCACCAAG 136850
Qy 1277 GTGTCACTTGTGAGAGGCTCGGAACCTCGGTGTACACGACGCGAGACTCTCTGAC 1336

Dh 136849 GCGCGCGCGCGCAAGGCTGTGGGACCTGCGCTACCGGATACGCCAACCCCGCAG 136790
Qy 1337 TGTGCTGATCATCAGCTTCTACGAGACGATGGGGAAGCTGTCTCGCTGGAGACATCG 13396
Dh 136789 TGTGGCTGGACCAACAGGCGCTGCGAGCTGACGCGCGCTGATCTACAACTGGATGG 136730
Qy 1397 TCGACGAGATGTTCGCGCGCGCATCTTGTGACGACATGCTGAAAGCTGCTGTTTTC 1456
Dh 136729 TCGATGCGCTGTTCACCGCGCGCGCTGCTGATGCTCATGTTGACGCTTCAACCGGATGC 136670
Qy 1457 TCGCGCGCTTAC-----TGAGAACATGGGGTGAACAGGTGCGCTGTTCCGCG 1510
Dh 136669 TAGAACCTTCGCGCGCGACGCGGATGCTGTGAGCGCTGCCCGCTGTGCGC 136610
Qy 1511 CTGCCAGCTAGAGCGCGGCGAGCGCAAAACGCGACCAACGCGCTGTGAGCGAGCAT 1570
Dh 136609 AAGCTCAGCGGAGTGTGCGCGCGCGCTTCAATGCGACGCGCGCGCTTCCCGAGCGCT 136550
Qy 1571 CGTGCACGCGCTGTTCGCGCGCGGTTGAGACAGCTGCCATGCACTGCGCGTGTGT 1630
Dh 136549 GCTTCGACAGCTGTTCGACGAGGAGCGCGCGACGCGCGCGCGCGCTGATTCG 136490
Qy 1631 CGCGCGCGAAGACCTACGATGAGAAAGCTTTGCGCGCTTCCGCGGACTTGGCGCGC 1690
Dh 136489 CCAACGACTTCACTGACCTTACGACCAACGCTGCGCGCTGTCCGCGCTTCCCAACG 136430
Qy 1691 GCGTCCGAGCAGAGGCGGCGCGCGCAACATTTGTCGCGGTGTGATGGAAGAGCT 1750
Dh 136429 GCTTCGCGCGAGCGCGGCTGTGAGCGCGGTGACCGCGGTGCGGTGTGATGCGCAAGGCG 136370
Qy 1751 GGGAGAGTGTTCGCGGTTCTCGCGGTCTGAGTCAAGCGCGCGCTTACGTCGATTCG 1810
Dh 136369 CGAGAGAGTGTGCGCGCTGCGCTGCGGATTCAGCGCGCGCGCGCGCTTATGTCGCGGTG 136310
Qy 1811 ATGCGGACCTACCGCGGAGGCTATTCACCTACTCTCTGATCATGTGAGGTAAAGCTCG 1870
Dh 136309 ATGCGAGAGCCCGCGCGCGCGCTGCGACGCACTCTGCAAGGACGCGCATCCGCGCTG 136250
Qy 1871 TGTGACGAGCCCATGTGCTGGAATGGCAACGTGATGCGCGCGGAGTCCAGCGGCTGC 1930
Dh 136249 GCTTGAAGAGGCGGCGATGCTGCTGCGGATCTGTGCGGAGCTGACCGAGCTTCAAGTGC 136190
Qy 1931 TCGTGAAGAGCGCGCGCTGCGAAGCGCGACGCGACG-----CTCCGATGATGCCA 1984
Dh 136189 GGTTCATGATGCTTCGCGAGCGAGCGGTCGCGGCTGCGCGGAGCGCGCGCGCGCGCTG 136130
Qy 1985 TTCAAGACCTTCCGATCTGCGGTATGATCATCTACCTTGGGATTCACAGGTTGCCA 2044
Dh 136129 CGCGCGATTCGCTGACCCGCGCTTACGTAATCTTACCTCGGCTTCAACCGCGCCCA 136070
Qy 2045 AGGGGTGATGATGATCATTCGCGGTGCGCGTCAACATCTGTGACATCAAGAGGCT 2104
Dh 136069 AAGGCTGTGATGACACCGCGCGCGCGGTCAACAGGTGTGATCAACCGCGCT 136010
Qy 2105 TCGAAATAGGCGCGCGAGACAGGCTGTGCGGCTCTCTCGCTGAGCTTGCATCTTCG 2164
Dh 136009 TCGGCAATACCGCGGAGATGTGTGTGCTTGGCTGTGCGCTTATCTTGAACCTGTGCG 135950
Qy 2165 TCTATGATGTGTTGGAATCTGCGCGCGCGGCTGAGTGTGTGCGCGAGCGCTTCA 2224
Dh 135949 TCTATGATCTTTCGGAAGCTTCCGCGCGGCGCGAGCTGTGTCTCCGAGCGCGCG 135890
Qy 2225 AGCTGGCGATTCGCGCGCATTTGGCGAAGTTGATCGAACGAGAAAGTGAACGTTGGA 2284
Dh 135889 GCGTGCAGATTCGCGCGCATCTGCTGCGCTTGGCGACGCGCATGAGGCTGACGTTGGA 135830
Qy 2285 ACTGCTGCGCGCGCTGATGCGGATGCTGTGAGCAATTTTGAAGGTGCGCGCGATTCG 2344
Dh 135829 ACTGCTGCGCGCGCTGCTGGAATGCTGCTGACGAGGCGGAGCGCGCGCGCGAGCGC 135770
Qy 2345 TCGCTAGGTCTTGTGCGGCTTTCGCTGTGAGCGCGCATGATTCGCGTGGGCTGTGCTG 2404
Dh 135769 TGGCGCGCATTCGCGCACGTCT---TCTGAGCGCGGAGTGGATCCCTTGCACCTGCGCG 135713

QY 2405 GCGAGCTCCAGGCCATCAGGCCCGCGTGTGCTGATTCAGCCTGGCGCGGCCACCGAAG 2464
 Db 135712 CCCGCTCTGGCGCCCTCTGGCGCGCGCGCGCTGATTCGCATGCGCGCGGCCACCGAAG 135653
 QY 2465 CGTCGATCTGTCGATCGGCTGACCCCGTAGGAAACGTCGACCTATCGTGGCGAGCATCC 2524
 Db 135652 CGTCGATCTGTCGATCGGCTGACCCCGTAGGAAACGTCGACCTATCGTGGCGAGCATCC 135593
 QY 2525 CCTACGGCGCGTCCGTCGCAACACGACGCTTCCACGTCGTCGATGAGGCGCTCGAAACCGC 2584
 Db 135592 CCTACGGCGTCCGTCGCAACACGACGCTTCCACGTCGTCGACACACACGTCGCGACT 135533
 QY 2585 GCCCGCTCTGAGTCCGGGCGCACTTACATTTGGCGGGGTGGGCTGCGACCTGGCTACT 2644
 Db 135532 GCCCGCGACACGTCACCGCGCGATCTCTGATCGGCGCGCATTCGCGCTCGGCTGACG 135473
 QY 2645 GCGCGGATGAGGAAGACCGCGAAGAGCTTCTGTCGTCGACCCCGAAGACCGGGAAGCGC 2704
 Db 135472 AGAACGATCCGAGCGCGACGCGCGCGCTTCTTCCACATCCGCTCGACGCGGAGCGCC 135413
 QY 2705 TCTAACAGACCGCGCGATCTGGCGCGCTACCTGCGCGCATGAGAAACATCGATGAGGCGC 2764
 Db 135412 TCTAACCGACCGCGCGACCTGGCGCGCTACCTGCGCGCATGAGCAACGCTCGATTCGCGGC 135353
 QY 2765 GTGAGGACCAACCAATCAAGCTTCCGCGATTCGCGCTGACGCTCGGGGAAATCGAGAAA 2824
 Db 135352 GCGCGCATCTTCCAGGTCAGATCGCGCGACCGCGATTCAGGTCGCGCGAATCGAAGCG 135293
 QY 2825 CGCTCAAGTCGATCCGGAAGTACGCGCGCGATTTGTCGCGCGGAAAGACGCGCG 2884
 Db 135292 CCTCTGTCGCGACCGACGAGCGCTGCGCGATGCGCTGTCGACCGCATCGCGCGCGCG 135233
 QY 2885 CGAACAGCTCCTTCTAGCTATGTGTCCCGAG 2919
 Db 135232 GCACACAGCGCTGCGCGCGATGCTGTGTCGCGAG 135198

RESULT 13
 AF516145 40156 bp DNA linear BCT 03-OCT-2002
 LOCUS Lynbha majuscula barbamide biosynthesis gene cluster, complete
 DEFINITION
 accession AF516145
 version AF516145.1 GI:23452290
 keywords
 source
 organism
 ORGANISM
 Lynbha majuscula
 Lynbha majuscula
 Bacteria; Cyanobacteria; Oscillatoriales; Lynbha.
 1 (bases 1 to 40156)
 Chang,Z., Platt,P., Gerwick,W.H., Nguyen,V.A., Willis,C.L. and
 Sherman,D.H.
 TITLE
 The barbamide biosynthetic gene cluster: a novel marine
 cyanobacterial system of mixed polyketide synthase
 (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an
 unusual trichloroethyl starter unit
 JOURNAL
 Gene 236 (1-2), 235-247 (2002)
 AUTHORS
 Chang,Z., Sherman,D.H. and Gerwick,W.H.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (28-MAY-2002) Microbiology, University of Minnesota, 421
 Delaware Street, SE, Minneapolis, MN 55455, USA
 FEATURES
 source
 1. 40156
 Location/Qualifiers
 /organism="Lynbha majuscula"
 /mol_type="genomic DNA"
 /db_xref="taxon:158786"
 5531..5818
 /gene="barA"
 /gene="barB"
 5531..5818
 /gene="barA"
 /function="probably involved in the activation of leucine
 to initiate barbamide biosynthesis"

gene
 CDS
 5531..5818
 /note="peptidyl carrier protein"
 /codon_start=1
 /transl_table=11
 /product="BarA"
 /protein_id="AA032974.1"
 /db_xref="GI:23452291"
 /translation="MSNSTLEPIDVVVPOTETEKOLAEIKWIVLSTIEIDRFMD
 IGENSLINVEVNVQTEKMGVSIKARPFDPDKSTIAELSEIDAVNQTH"
 5850..6752
 /gene="barB1"
 5850..6752
 /gene="barB1"
 /function="probably related to chlorination"
 /note="similar to the syringomycin Syb2 protein and the
 coronamic acid CmaB protein"
 /codon_start=1
 /transl_table=11
 /product="BarB1"
 /protein_id="AA032975.1"
 /db_xref="GI:23452292"
 /translation="WKTAKAELSYIVSVEKQFPDNGYIGPTLPOPEITLAME
 IRMDLDRETPAPFNSSKMYDRHLDIALSDIISHPKIAHVSLLGSNVLCMTYAP
 PKYDDEGTDMHQALESFVEFEKSLVPTESHDGPMELTAWAFSEATRENGCMKMP
 GTHNTWYDEHRIEPEPDKINOKLTGCKTGVYDYDKLDPNMEPDSQAVHLE
 MEGQFLFTSRCHMGSEPNSSGSIKRGMSIRVSTVDPVRYVPMDSFEHFEKFMLE
 RYATVYAGEDTYKHKIRPLQ"
 6812..7705
 /gene="barB2"
 6812..7705
 /gene="barB2"
 /function="probably related to chlorination"
 /note="similar to BarB1, syringomycin Syb2 protein and
 the coronamic acid CmaB protein"
 /codon_start=1
 /transl_table=11
 /product="BarB2"
 /protein_id="AA032976.1"
 /db_xref="GI:23452293"
 /translation="WKTITAEPRYLYSVEKQFQENGYIGPTLFPPEMELMYGI
 KMDLLDRETPAPFNSSKMYDRHLDI KALNDIISHPKIAHVSLLGSNVLCMTYAP
 KYPDDEGTDMHQALESFVEFEKSLVPTESHDGPMELTAWAFSEATRENGCMKMP
 THTTWYDEHRIEPEPDKINOKLTGCKSGFYDYDKLDPNMEPDSQAVHLE
 EPGQFLFTSRCHMGSEPNSSGSIKRGMSIRVSTVDPVRYVPMDSFEHFEKFMLE
 YSTVYVAGEDTYKHKIRPL"
 7859..8665
 /gene="barC"
 7859..8665
 /gene="barC"
 /function="proofreading or editing role or chlorination
 role"
 /note="thioesterase II"
 /codon_start=1
 /transl_table=11
 /product="BarC"
 /protein_id="AA032977.1"
 /db_xref="GI:23452294"
 /translation="WQTDSSWLFDETPKCEASKMLFCFHYAGAGSITQDMSRAL
 PDEVELVAIQLPGRKRCYEPMSNIEDINVLQOESPFISKPAFEGHMGALIAF
 DLIRLRQOELGQOFLWVGFDAPOLAPKSSNMHQLPDEPQALINDYQILPQVIA
 ONELFELLRLRSDIKLGEITYIKOBEPLDCRIIALYSTEDSFVSKSGMYSQEQT
 KGEFLVTPPDHFLHSAELVNTICTELQRYIKANDNCSTKVALQKS"
 8700..10253
 /gene="barD"
 8700..10253
 /gene="barD"
 /function="adenylation; involved in activation of leucine"
 /note="sequence similar to adenylation domains of NRPS"
 /codon_start=1
 /transl_table=11
 /product="BarD"
 /protein_id="AA032978.1"
 /db_xref="GI:23452295"
 /translation="WGVYIANLVTRHSKSPHKCAITDGTLELYGVWEASNPARK

LIENKQKGRVNVLAISKQTLIMVIGLIGFACGICIVHEDLMDKMPVNRINKITITELIPS
VITLDOELHLIVSENVPOEKIVILEQHLIPITSTVEKDLRLOIESIEDIAYCIYTS
GSTEDPKVLILKSENVIDFEGRETFYVDVEDSRVCSFSPHDPASVIDILPELYOGA
MLYLYSDVVLHLELFEVINKNSVTHFAFGSMGLIAQTKFEDVILIKDLITLITGAD
VPDIKVIQKMLKGENIKITIKINGCTBETCAATITIDRPNRELYPTIGKPLKNAK
LYLVDOQOEIYDEPIPEGLIGTQVAGVWNRQSKERLVEKGTCTKTDICK
ALAMLANETSKDSALKQIKENHNOELPQVMEVROQIVMEKFPQLSSGKTRKTLSSILE
RDLASKVH"
10313..15433
/gene="bars"
10313..15433
/product="bars"
/function="involved in activation of tricholeucine and
condensation with malonyl group"
/note="NRPS; A, PCP, KS and AT domains are located on
BarE"
/codon_start=1
/transl_table=1
/product="BarE"
/protein_id="AAN32979.1"
/db_xref="GI:23452296"

/translation="MKNDPKGTHQKNDYEVNHSKSTSGSLNTPNSPTLCLDLPR
AALNKGITVQADGSEYOQSYODLLDAKKILAGRWMDLKPQDKVILQLEBHDPI
PAFWGCLLGFI PVPLSIAPTYOQANVTIKLHNAMQMLHEPLILITNRELPEIRLS
THNLIDFQGTVIDELRKQDPTQOIEYRGQPELILALLTSGSTLPGKVMHSHNLS
ESAGTIQKMHFDEAVTLNMMPLDHVGGIIMLNVAVALGCOOVAPEKYLQNLPH
KLIMRQGLAETVLPAPFGMSTSGITTDPSISNTSDMPFVGLGPIYACLRILY
NDNDEIVHEGAIQKLVQKASVTSGYONPEINQAFADMPFGDGLYGLACRLIV
TEGRNDVILINGVYPAHEIAVEEVGVNISPTAACALIKTAKISSSDOLAIFSA
VENIETIPEVLKGLIKKIRGAVVQNIENIDYLPVAKESIKPTAKISIKOROLK
FEAGEFDLITVGGSLRIMGSOETALAPOTEVEQVAVNIWODVGLGDKIADGNF
ELGSHSLILITLOLOSFGKELSIAMFHPITELAOVPTQOPEDSAACGYAR
AQIOEVRKADSTELAVIAGMAGRPFGADNLDQFQNLONGESTITFFSDSLASGD
PKLSNPYVRAVPSILSDVEYFDAHFEYTRAEFMDPQRLMECAMESELADAGD
PLTYSIGLVAAGMLSTYFTHNLNPNDRDEHDLKVLTPDSMGGLQVLANDKV
LITRISHKLANLTPGSVNIOTACSTSLVTVRAVOSLNGECDLALAGSVQVORVG
HLHOGELIASPDGCRAPDAQAEFTIPNGVGLVAKRLEBALDDGHIYAVIGSAL
NDNSNVAYRAPNGEOTTVAAEIAANINPITITVEAHGCTILGDIIEGALS
QAFQAKTORQPCALGSKVIVNGHQLTISVGLIKTVLALYHKIAPSLHPOTPNK
IDFANSPFVNTMLSEWERNHPRRAGNSVIGFTGNVAILLEAPENAOIKPTENEDG
AKLEDRSHTLTLAKTAPALQALVKRQRLSDNTELEDIDICTACVAGAEHRL
AIVSDSIDLECKLANVLDNAELSPBOTSLAQOCTASVRAQIALFLTQCGOYL
MGRLOTOSVFRDAIDCCQIILQSELEVSILDLXPOTANAAARNMTAATPOPL
PAIEYALAQWMSGCIOPAYVMGHSGEVVAATAGVDTGKLIARRLAQOLP
AGXAMVWASBEERVSILSPHTEKISIAIANGPSVVISGVAELGAIISQLESAGV
KTKSLQVSHAFHSPMEPIIAEFTVAQQLYHSPRIPLISNVGVKADQEIARAQV
VNHRIQVRFAPHSMBALSEFGDCLFIEIGRPILILGMROCLPSTGVMWFLSLATVD
DEQOMLSIAGLYQGVVWMSLNOEYSYQVTLPTYPKRYMIEKYETNOKPT
SATSNOKNNTIFRLINQNVLDIOLYIKKTDLSREBEVKLSRPAQILINLVHQBNE
LITNKDKERKIDNMQLQMDSENKILSLSEKPSQNS"
15476..16990
/gene="barF"
15476..16990
/product="barF"
/function="involved in addition of O methyl group"
/note="contains methyltransferase (MT) and ACP domains"
/codon_start=1
/transl_table=1
/product="BarF"
/protein_id="AAN32980.1"
/db_xref="GI:23452297"

/translation="MSDLEMMNNLQENLSTLSPBOROFEFKTIAENKPHQSLGQER
IAEKSTBERLSRYRALDHSNTADVARPFSBEVDFWMLDHPDHPQOTSL
SLQAQAEKVALFRGIDFSSIKYWDIGCGSHLIDATLVHVLQLDGVNISPEVYK
AGBOKITOGISYRILYINRDSAKOPLDEYDLIFSCGVINHILKKEBEVFNISQAN
DSGEFFVAEISNPLPIYIDAKSAAYVTRSKMAOLLNARKVEVDNFANIGIYAN
LYDENFDNMLARLCOGYDENAKDHPHLLGELLKRLAYLILITQOKSFTKERT
LRNOEKONTLPYSKITEKNSPKPKVLLPPFRSGTTPSQGREDRPOPLSODLS
DAQVGLAFYLTVAAILKETQAEINETSITFVGVDSLWELBENRQADLVADISV
ADLMGDNISRLAIKVBELLHRNAQADLESKMDGDNLOEIALODKKMRBGRLL"
16987..25650

/gene="barG"
16987..25650
/product="barG"
/function="responsible for the recognition and activation
of phenylalanine and cysteine and for the release of the
barbamide intermediate from the assembly line"
/note="NRPS; two NRPS modules are located on the protein
containing C, A, MT, PCP, Cy, A, PCP and Te domains"
/codon_start=1
/transl_table=1
/product="BarG"
/protein_id="AAN32981.1"
/db_xref="GI:23452298"

/translation="MNLSEFLQELVLSGMQFAEEGVCFQAPPADSTQVLAOLKQH
KRDLITLQHEPVLQVPLGQGGIIMPLQLEPDNPVNSVATRIYSQVNTTQW
OTPEAKRHRHLICSTPKGERTPIRQHSBDDIVQDASWDNENLOTOVAAHNRH
FDLQTDPMVRVMTFRSEQEHILITLHIAIMDSANATVLELSLQAHQACAVAD
LPSLQHTYQDVVWQQLVEGSKESLMTWQQLAGELPVNLPTBRPPIQTNRG
AVYRFQLEHVLTVQKALQABGATLWTLAOLVILHRTVGEQDLVGSPTSGTR
PEFTSVGVFVDSMWBARQVSGSLSPRELFQVQVITDALADQDYPSSLVEKLOPE
RDLRSRPIQVFPGLHNPLOSETQQLFLGETKTLVHMGMEVFELFQYESLBDVL

Query Match 14.4%; Score 609.4; DB 1; Length 40156;
Best Local Similarity 52.9%; Pred. No. 2e-86;
Matches 1499; Conservative 0; Mismatches 1291; Indels 45; Gaps 7;

QY	179	CCATCGTCCCGCCCGGACGAGGCGACGCTCGGTTTCTCTGACAGACATCCAGAT	238
DB	21581	CAATTGTCCTGCGCCCAAGCGGCTTACACCTTTCCCTCACCGAGATTCACAAAG	21640
QY	239	CCTACTGCTGGGCGGACAGAGCGTTTACGCTCCCGAGCGGATC---ACGCGTATC	295
DB	21641	CTTATGCTGCTGGCGGCAATTCGACATTGACTGGGTAATATACACGCGATGTTATC	21700
QY	296	GCGAATACGACTGACGATTCGACGTCGCGAGCGTACGCGGCTTTCGAAAGTCG	355
DB	21701	TGGAACTGCACTGGAGAAATTTACCTTTAGACCCGTTAAGTCAGAGCTTGGCAGAAAT	21760
QY	356	TGCGCGCGGACGATGCTTCCGCGCCACACGCTGCCGACATGATGAGGTATCGAGC	415
DB	21761	TAGACCATCAAGATGCTGGGAATGATGATTCAACCAATGGTGAACAGAAAGCTTTCG	21820
QY	416	CTAAGTCGACGCC---GACATCGATCATCATGCTGCGGCGCTCGACCGGACACAC	472
DB	21821	AACAGTTTACCCCTATCAAGATTGAAGTTTAACTTCCGGGTCAACCAAGAAATCG	21880
QY	473	GGAAGCGAGGCTGCTGCTGTCGGAAGTGCATGTCGCACCGCATTTATGACACCGAGC	532
DB	21881	TCTCAACCGAATCAAGAAACCATTTGTTATCGCTGTCACAGAGATGTTCCCGCTGGAG	21940
QY	533	GCCCTCGCTCTATCAGCTGCTGCCGCTGGCGTGCAGAGCGGCAACCGGCTCTGTCG	592
DB	21941	AGTGCCCATATTAAAGATTCGTGACTGCTGAGCGAGATCAACGCTTACACT	22000
QY	593	TCAGTATCGATCTGATTAAAGCTTACCTAGGACGCTGCTCCATCATCTTCAAGATGCG	652
DB	22001	GGACCTTGATGCGCTTAAATGCTATGCTGAGACATGATCATGTTGTTGGCAAGATGCG	22060
QY	653	TCAGCTTTCAGAGATCCGAGACTCTCTCTGCTGCTGTCGAGACTCTGTAACGCGATT	712
DB	22061	TGCAGCTATACCAAGAACCCAGACAGTTTCCATCCAAATATGACTCATATTCGGGACT	22120
QY	713	ATGACTTCGCGCTGAGTCTTCGGAAGAGTGTAGGCGCATCAAGATCGATGATTAAT	772
DB	22121	ATGTGTTAGCAAGACTAGTCTCAAGACAGCGCCCAATATCGACGCTTCAACAGATTT	22180
QY	773	GGAAGCGGCGATGCGGAGCTCCACATCGCGGACGCTTCCATGAAGCGCATCAT	832
DB	22181	GGTGGATTCGTTAGACACATCCCCCTGACCAAGAACTTCCCTTGCTCAAGCAATCG	22240
QY	833	CTACCTCGAAGAGATCCGCTTCGCGACACAGGACATGCGTCCGTCGATCTCTGGG	892
DB	22241	CCACACTAGAGCACTGAAATTAATGTTATTCGGGCTGAACTGCTGCCCCAGATGGC	22300

QY 893 GTTCGATTGAAGCGGCTGTTCGGGAGCGGGGCTGACCCCGACGGGCGCTGATCTGGCTG 952
Db 22301 AGCAATTTGCAACCCAGAGGAAAAACAAGCAGATTGACCGTCAAGGGGCTTGTGAGCGG 22360
QY 953 CATTTTCCGAGGTGATCGGCGCTGGAGCGGAGCCCGGGTTTACGTCAACATTAACG 1012
Db 22361 CATTTCGCGATTTCTTACGCTTATGAGGCAAAAAGTCCCAATTACGATCAATCTCACCC 22420
QY 1013 TTTTCAACCGGCTCCCGTTCATTCGCGCGTGAACGATATCACCGGGACCTTACGTCGA 1072
Db 22421 TCTTTTAAACGGCTACCTTTGCAATCCCAAGTCATATATTAGTGGAGACTTCACTTCCC 22480
QY 1073 TGTCTCTCTGACATTCGACCAACCACTCGGACCAAGAGCTTCCAAAGCGCGCTAAAGGCTTA 1132
Db 22481 TGAACCTTTAGAGGTCAATCAAAAAAGCAGCTCCCTTGTCTGACGCTGCTCAACGCTC 22540
QY 1133 TTCAAGAGCAGCTGTGGAGGAGATGATCACTGCGACGTAAGCGGATATCGAGTCCAGC 1192
Db 22541 TCCAGGGAACAATTGGCAAGACTAGACCAACGCTATGTGGGGGCGTTGAGTCCAC 22600
QY 1193 GAGA---GGCGCCCGGCTCTGGGGATCCAAAGAGCGCATTTGTTCCCGGTGCTC 1248
Db 22601 GAGACTTTCGCGCTCAGCGCGGATGATCAACCCATGGGGGTTGATTTTACAGATACGC 22660
QY 1249 AGGAGCGGCTTAACAGAGAGTCGTTGGTTCACCTCGTTGAGAG----- 1295
Db 22661 TGGCGCTGAATAGTCGAGAGAGAGAGATGCGCTGCTCAATGATGACGCTTGCCAT 22720
QY 1296 -----GCTCGAATCCGCGTGTACACAGCAGCAGACCTCTCAGCTGCTGCTGATC 1348
Db 22721 TGGACCAATTTGGAGAAACGTGTTATATGCTCTTAAACCCCTCAAGTGTGCTAGACA 22780
QY 1349 ATCAGCTTAACGACGATGGGAGCTTGTCTCGCGTGGACATGTCGACGAGTGT 1408
Db 22781 ATTCTGTTCTGAACAAATGAGTGTCTTACTCTTAATCTGAAATGATGCAAGACTCTC 22840
QY 1409 TCCCGCCGACCTTTCGAGCAGATGCTGAAGCGATAGCTGTTTTCGCGCGGCT-- 1466
Db 22841 TCCAGAGGGGTTCTCAATGATGATTTCAACAGCTACTACATGTTGGCAACAGCTCG 22900
QY 1467 ----CACTGAGAACCATGGGGGTGAACAGTGGGCTGTTGCTTCCGCTGCCAGCTAG 1522
Db 22901 CAACCTCGATGTGGCTTGGGGCGACACTTGGCCCTCAACTACGCCCCCTCTCTCACTTA 22960
QY 1523 AAGCGCGGAGCGGCAACGCGACCAACGCGCTGCTGAGCGAGCATACGCTGACGCGC 1582
Db 22961 CCCAACGCTCAAGTTAAATGAACCTATGCCCCCTGTTCTGMAAGAGACCTTCATATTT 23020
QY 1583 TGTTCGCGGGGGGTCGAGCGAGCTGCCATGAGCTCGCGGTGCTGGGGCGGCAAGA 1642
Db 23021 TGTTCGAGGCAAGTTTCAGCGAGCAGACAGAGCGGATCGCCCTCATACCCCGCAACGCA 23080
QY 1643 CGCTACGTACGAAAGACTTTCGCGCGTTCGCGGCGACTTGGCGCGCGGCTGCGGAGC 1702
Db 23081 CCTCACTCACTAATGAATCTTACACGAGCTCAGCGGCTCGGCGCAAGTTCAACAT 23140
QY 1703 AAGGGGACGCGCGAACAATGCTGCGGTGATGAGAAAGGCTGGAGCAGGTTG 1762
Db 23141 TGGAGACCAACCCCAACACCTTATGACGCTGCTCATGGAATAAGGCTGGAGCAATATG 23200
QY 1763 TGGCGGTTTCGCGGCTGCTGAGTCAAGGCGGCGCTTACGTCGCGGATGATCCGCACTAC 1822
Db 23201 TGTCTCTTAAAGCACTTTTGAATGCTGTGACCTTATCTCCCATTTGATGCGCGCTCG 23260
QY 1823 CGCGGAGCGTATCACTACTCTCGATCATGTAGGTAAGTCTGCTGACCGCAGC 1882
Db 23261 CCCAGAGGCTCATATGCTCTTCTGAGAGAGAAAGTCAAACTGCTGTGACACAG 23320
QY 1883 CATGCTGATGCAAACTGTCAATGCGCGCGGGAATCCAGCGGCTGCTGTGAGCGAG 1942
Db 23321 CTGCGCTAAACGCTTATGATGCTGACAGACCACTGCACTGCTGTGAGT-----TG 23374

QY 1943 CGGCGTGAAGAGCGGACCGGACCAAGCCTCCGATGATGCCATTCAGACACTTCGATC 2002
Db 23375 CTTCCCAACACAGAGATCATAGACATCCCCCTGAGGCAAACTTTAGCAGACGATTT 23434
QY 2003 TCGGATATGCTATCTTACACTTCGAGATCAAGGTTGCCCAAGGGGATGATGATC 2062
Db 23435 TAGCTATGATGATCTTTTACTTCGCTGCTTCAAGAAACCCCAAGGGGATGATGATC 23494
QY 2063 ATCGGGGTGCGCTCAACACATCTGAGATCAACAGACGCTTGCAAAATAGGCCCCGAG 2122
Db 23495 ACCGAGAGGCGGTTTAAACACATCCAAAGATATTAAACAAAGGTTTACGTTACGCCACTG 23554
QY 2123 ACAGGATGCTGCGCTCTCTGCTGAGCTTCACTCTCGGTCTATGATGTTGGGA 2182
Db 23555 ACCGATGCTGCGGCTCTCTGCTTCAACTTTGACTTGTCTTACGATATCTTTGGCC 23614
QY 2183 TCTTCGCGCGGCGGCTGATCGATGCTGTGCGGAGCGCTCAAGCTGCGGATTCGCGCGC 2242
Db 23615 TCTGCGAGCGGGGGGACCCCTAGTCATGCCCACTCCAGAGGACCCAAAGACCCAGTCC 23674
QY 2243 ATTTGGCAGATTTGATCCAAAGAGAAAGTGAACGCTGTGAACTCGGTGCGGCGCTGA 2302
Db 23675 ATTTGGTGAATTAATGAACCAACCATCACTCTGGAACAATGCTCCAGCTGA 23734
QY 2303 TGGGATGCTGCTGAGCATTTTGAAGGTGCGGCCGATTCGCTGCTAGTCTGCGGC 2362
Db 23735 TGCATATGCTATGAAATATCTGTGAGACACCAAGACCAAGTATCTAGAGACTTCGAC 23794
QY 2363 TTTTCCTGTGACCGGCGATGATCCGCTGGGCTGCTGCGAGCTCCAGGCAATCA 2422
Db 23795 TTTGCTGTGATGATGATTTGATTTCACTCAATTTACCAACCAAAATCCAAATCATTAT 23854
QY 2423 GGGCGGGGCTGCTGATCAAGCTGGGCGGGGCAACGAAAGCTGATCTGTGCTCATC 2482
Db 23855 GGGCCCAAGAGAGGTGTGCTGAGGAGGCAACAGAGCTTCAATTTGCTGTGT 23914
QY 2483 GGTACCCCGTGAAGAACGTCGACTATGTGGGCGAGATCCCTCAAGCGCGCTGCTGC 2542
Db 23915 ACTATCCATCTACCTGATGAGCCGGAATGGAAGAGATTCCTTACGCAAACTCTAG 23974
QY 2543 GCACACAGCGTTTCAGCTGCTGATGAGCGGCTGAAACCGCGCTGCTGGGTTCCGG 2602
Db 23975 TGAATCAAAAGCTTATGCTCAATCAATCTGACCTTGTCTTATTTGGTTCCAG 24034
QY 2603 GCGAATCTACATTTGGCGGGGCTGCGCTGAGCTATGAGGCGATGAAGAGAA 2662
Db 24035 GTACGCTTCAATTTGGTGTGATTTGGCTGCGGAGCTATTTGGCGGATGCAAAAAA 24094
QY 2663 CGCGAAGGCTTCTCTGTCACCCCGAGACCGGAGAGCGCTCTTCAAGACCGGCGATC 2722
Db 24095 CGAATGCCAGCTTATCTGCAATCCCAAGCAACAGTAAATTTTGGGTGAGGATTTTCCAAAGTCA 24214
QY 2723 TGGGCGCTACCTGCGGATGGAACATGAGTTATGAGGCGTGAAGCAACCAATCA 2782
Db 24155 TGGCGCGCTACTTACCCAGCGCAACAGTAAATTTTGGGTGAGGATTTTCCAAAGTCA 24214
QY 2783 AGCTTCGCGGATACCGCGTGTGCTCGGGGAAATGAGAGAAACGCTCAAGTGCATCCGA 2842
Db 24215 AATATGATGATGATGAAATGAATCGGTGAGATTTGAGCAACCTGCTGGGGAATGCTA 24274
QY 2843 AGTACGCGACCGGATGATTTGCGCGCTGCGGAAACGAGCGGCGAACAAGTCTTCTAG 2902
Db 24275 CAGTTTAAAGAAACG---TTGTGCTGCTGTGTGGAATCAAGATTAAGCAGTTGTGCTG 24331
QY 2903 CCTATGTGTTCCCGAGGCAACGAGAACGCGCTGCGAGCAGAACGCACTTCACGA 2962
Db 24332 CCTATGTGTTTCACTCAAGATCATCTGACTGCAACAGAAAGCTTCACGATG 24391
QY 2963 CCGAGCGATGACG 2977
Db 24392 ACATGCGCATGACG 24406

RESULT 14	SC039132	302007 bp	DNA	linear	BCT 11-FEB-2003
LOCUS	SC039132				
DEFINITION	Streptomyces coelicolor A3(2) complete genome; segment 29/29.				
ACCESSION	AL339132	AL335575	AL391338	AL391395	AL391454
	AL450165	AL450222	AL450289	AL450350	AL450432
	AL596144	AL645882			AL512944
VERSION	AL39132.1	GI:24413926			
SOURCE	Streptomyces coelicolor A3(2)				
ORGANISM	Streptomyces coelicolor A3(2)				
	Bacteria; Actinobacteria; Actinobacteriales; Streptomyces				
REFERENCE	1				
AUTHORS	Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wierzbicki, A., Woodward, J., Barrett, B.G., Parkhill, J. and Hopwood, D.A.				
TITLE	Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)				
JOURNAL	Nature 417 (6885), 141-147 (2002)				
MEDLINE	21996410				
PUBMED	12000953				
REFERENCE	2 (bases 1 to 302007)				
AUTHORS	Bentley, S.D.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk				
COMMENT	On or before Oct 26, 2002 this sequence version replaced g1:20520855, g1:9843813, g1:9857138, g1:9857140, g1:20520741, g1:20520787, g1:20520787, g1:20520842, g1:20520681, g1:20520707, g1:20520788, g1:20520682, g1:20520861, g1:20520876.				
FEATURES	Location/Qualifiers				
SOURCE	1. 302007				
	/organism="Streptomyces coelicolor A3(2)"				
	/mol_type="genomic DNA"				
	/strain="A3(2)"				
	/db_xref="taxon:100226"				
	116. 1390				
gene	/gene="SC07543"				
	/note="synonym: SC8G12.19"				
CDS	116. 1390				
	/gene="SC07543"				
	/note="SC8G12.19, possible Crp-family transcriptional regulator, len: 424 aa. Similar to several including: Bacillus subtilis SW.XYLR_BACSU (EMBL:W27248) XylXose repressor (384 aa), fasta scores opt: 339 z-score: 347.1 E(): 7.1e-12 23.3% identity in 390 aa overlap and Streptomyces coelicolor TR.050502 (EMBL:AL009199) probable transcriptional repressor protein, SC7B7.05 (403 aa), fasta scores opt: 691 z-score: 697.8 E(): 2.1e-11 33.9% identity in 387 aa overlap. Contains a Prosite hit to PS01125 ROK family signature and Pfam hits to PF00480 ROK, ROK family with the putative helix-turn-helix motif situated between residues 32..53 (+4.61 SD)."				
	/codon_start=1				
	/transl_table=11				
	/product="putative Crp-family transcriptional regulator"				
	/protein_id="CAB90901.1"				
	/db_xref="GI:7799279"				
	/translation="MTGQMPSGVATVRRTHEERVKALQENGMSRGEIARVVGLS RTTSEIGNLNGAIVVDTSRRSGRPARLADPSAQFMVDGHRVHV VVADAHEIMSVSVREDDTATPTPTREALEIYDRLSGNGYHGAOGIGICVPGP YPARBGAMPATPCTTTLRPAPECDVAFARFPAPVIVNNTRLALAAISGADS VADIVYLVSDVGGLVGGQLVVGGSLGELGAVVEBAGRPCRCGKGCLETTA SVPGTLAACWBFGLRLENTIDDLAAVRRAHPVDVRLREAAGALGRVVGAATMLNPA				
	KVIVGSEITRLAPLVVEQVAATLAELIFPTASAGPVVAARLSDDDGAGALAAVHSS SPLIARYPEPTADVGRSDAPHSATGAAHVRP"				
	206..298				
misc_feature	/gene="SC07543"				
	/note="Pfam match to entry PF000325 crp, Bacterial regulatory proteins, crp family, score 15.80, E-value 0.0014"				
misc_feature	680..958				
	/gene="SC07543"				
	/note="Pfam match to entry PF00480 ROK, ROK family, score 96.80, E-value 1.1e-27"				
misc_feature	794..877				
	/gene="SC07543"				
	/note="PS01125 ROK family signature"				
gene	1485..2288				
	/note="SC07544"				
	/note="synonym: SC8G12.20"				
CDS	1485..2288				
	/gene="SC07544"				
	/note="SC07544"				
	/note="SC8G12.20, possible ABC-transport protein, membrane component, len: 267 aa. Similar to many other transporters e.g. Escherichia coli SW:TAUC_ECOLI(EMBL:D85613) taurine transport permease protein, Tauc (275 aa), fasta scores opt: 547 z-score: 643.1 E(): 2.3e-28 37.5% identity in 248 aa overlap. Contains multiple possible membrane spanning hydrophobic domains and a Pfam match to entry PF00528 BPD transp. Binding-protein-dependent transport systems inner membrane component."				
	/codon_start=1				
	/transl_table=11				
	/product="putative ABC-transport protein, membrane component."				
	/protein_id="CAB90902.1"				
	/db_xref="GI:7799280"				
	/translation="MAADTSAQAESAPEPSRSGRRGLPVYANVSVALGATLMALA SAGKRLTPPEVDVRASTLIGDGLADALASTLRVILGELGAVAAVPGFLMGWIG LKRLRIPMWIOFPTTIPPLAIIPYAVVAGIDETPKLFIPLAFLACVSTFGQVNV VDRFLINAAVLGAGDATTIFARVVVPASTPTIIVGMVGLSAAVATVAAELIAOEG LGYRMONAOLYDPTIFVGLISIGILMDRVLTLAERRLTRWQERR"				
	1953..2168				
misc_feature	/gene="SC07544"				
	/note="Pfam match to entry PF00528 BPD transp, Binding-protein-dependent transport systems inner membrane component, score 44.90, E-value 1.8e-09"				
gene	2285..3097				
	/gene="SC07545"				
	/note="synonym: SC8G12.21"				
CDS	2285..3097				
	/gene="SC07545"				
	/note="SC8G12.21, possible ABC-transport protein, ATP-binding component, len: 270 aa. Similar to multiple ABC-transport proteins e.g. Escherichia coli SW:TAUC_ECOLI(EMBL:D85613) taurine ATP-binding protein, Tauc (255 aa), fasta scores opt: 665 z-score: 772.8 E(): 0.43e-4% identity in 235 aa overlap. Contains a Prosite hit to PS00017 ATP/GTP-binding site motif A (P-loop) and a Pfam match to entry PF00005 ABC_tran, ABC transporter."				
	/codon_start=1				
	/transl_table=11				
	/product="putative ABC-transport protein, ATP-binding component."				
	/protein_id="CAB90903.1"				
	/db_xref="GI:7799281"				
	/translation="WISTPAKISVQHYTKTFSGLREFTALDVSLDIANHEPTVVG PSQGGKSTLNNILAGLAPTSGRALVGVAPSGPGRGVFOQYAFPMVLTQONVE FGILRTTGVPRERRARAHEPTELVGSRFADALPTKISGGRCRCATARAADVPSIL LMDPPFALALTEVTKQEOELIKTWSQSDRTVLEITDVDEAVFLARVVMAAPGR VYVDVDDPPIIEREARFLSPEFALANRRWRSYVHGEGRALASGHSPP"				
	2393..2929				
misc_feature	/gene="SC07545"				
	/note="Pfam match to entry PF00005 ABC_tran, ABC transporter, score 224.10, E-value 2e-63"				

```
misc_feature      2414..2437
                  /gene="SCO7545"
                  /note="P500017 ATP/GTP-binding site motif A (P-loop)"
gene              3094..4116
                  /gene="SCO7546"
                  /note="synonym: SC8G12.22"
                  /gene="SCO7546"
                  /note="SC8G12.22, possible solute-binding lipoprotein,
                  len: 340 aa. Similar to several e.g. Escherichia coli
                  SW:TAUA_ECOLI (EMBL:D85613) taurine-binding periplasmic
                  protein precursor, Taura (320 aa), fasta scores opt: 235
                  z-score: 267.4 E(): 1.9e-07 23.3% identity in 287 aa
                  overlap. Contains a possible N-terminal signal sequence
                  and an appropriately positioned Prosite hit to P500013
                  Prokaryotic membrane lipoprotein lipid attachment site."
                  /codon_start=1
                  /transl_table=11
                  /product="putative lipoprotein."
                  /protein_id="CAB90904.1"
                  /db_xref="GI:779282"
                  /translation="MTRKGPATMSPVLTFRSLTTTLAAGVAVLLAACGSSSGSSD
                  /KRPFGYINDFNASLAIIEAGLWKHGLSADAVFTNGPIQIQLGTDNDPFGIG
                  PCAMVLPSAGQAKVVAINTGNSDRVLAQPGITSMQQLKATVAVEGTSGMISLA
                  LEKAGMTKEDLVPMIDPESTIVASFOODGAFMYPAARATTKKQVPLVLANAD
                  FEQDVSFPAFVAGNKVAVDQPEKTKVLAIVREANTPESBETDAIKUTAKKATPV
                  EGVKADANNKVLVAELDELTRDGTIVAMLEGMDYFVNAKGLKKPADPKTYITGLD
                  FTGAGK"
misc_feature      3163..3195
                  /gene="SCO7546"
                  /note="P500013 Prokaryotic membrane lipoprotein lipid
                  attachment site"
                  /gene="SCO7547"
                  /note="synonym: SC5F1.01, SC8G12.23"
                  /gene="SCO7547"
                  /note="SC5F1.01, possible sulfatase (fragment), len: >136
                  aa, low similarity to SW:YIDU_ECOLI (EMBL:L10328)
                  Escherichia coli hypothetical 57.3 kDa protein in
                  Emdr-Glyc intergenic region putative sulfatase YdJ1 (EC
                  3.1.6.-), 497 aa; fasta scores: opt: 147 z-score: 185.4
                  E(): 0.0085; 29.6% identity in 115 aa overlap
                  SC8G12.23, possible sulphatase, partial CDS, len: >330 aa.
                  Similar to many Eukaryotic sulphatases e.g. Homo sapiens
                  precursor (EC 3.1.6.13) (550 aa), fasta scores opt: 326
                  z-score: 382.2 E(): 7.9e-14 28.4% identity in 423 aa
                  overlap. Also similar to Escherichia coli
                  TP:YIDU_ECOLI (EMBL:L10328) hypothetical 57.3 kd protein
                  (497 aa), fasta scores opt: 221 z-score: 0.0 E(): 0.0
                  31.2% identity in 397 aa overlap. Contains a prosite hit
                  to P500523 Sulfatases signature 1 and a Pfam match to
                  entry PF00884 Sulfatase, Sulfatase."
                  /codon_start=1
                  /transl_table=11
Query Match      14.2%; Score 600; DB 1; Length 302007;
Best Local Similarity 52.8%; Pred. No. 3.9e-85;
Matches 1622; Conservative 0; Mismatches 1350; Indels 102; Gaps 11;
```

```
QY      182 TCGTCCGCGCCCGAGCCGAGCGGAGCGCTCCGCTTCTCTCAAGACATCCAAATCCT 241
DB      147644 TCGGCGCCGAGCCCGAGCACCGGACGAACTTCTCCCTTACGACATCCAGCCGCT 147703
QY      242 ACTGCTGAGCGCGGACAGAGAGGTTTAC---GATCCGAGCGGATTCACCGCTTATCCG 298
DB      147704 ACCCTAGCGCGGACCGACCGCTACGCTACGGGGGGGCTGCGCTGACCGCTACGTG 147763
QY      239 AATACGACTGATACGATATCTTCAACCTGCGGAGCTGAGCCGCTTTTGGAAAGTCGTG 358
DB      147764 AGCTGCGATATCCCACTTCAACCTGACAGATACCTCGTGTGGGCGCAATCGTTC 147823
QY      359 CGCGGACAGACATGCTTGGGCGCCACAGCTGACCGCATATGAGGATGATGAGGCTTA 418
DB      147824 AGCGGACAGACATGCTTGGGCGCCATCACTCAACAGAGGCGCACAGCGGCTTCCG 147883
QY      419 AAGTCGAGC---CGACATCGAGATCATGATCTTGGCGGGGCTGACCGGAGCACAGG 475
DB      147884 AGTGCCTCGCGCTGTCCGCGGAGACGACAGACTGACCGCTGCGGCGCCGACGCG 147943
QY      476 AAGCAGGCTGCTGTGTCTTGGAGATGTGCGACCCGATCTATGACACCGAGCGCC 535
DB      147944 AGGCAACGATGCGGACACCCGCGCGCGGCTGTCCGCGAGCGCCACACGACAT 148003
QY      536 CTCGCTCTATCAGCTGCTGCGCGCTTGGAGTGAAGAGCGGCAACCCGCTCTGCTCA 595
DB      148004 GGCCTCTT---TCACCTACCTGACGACCCGACCGACCGCGGCGCTATCGACCTCT 148060
QY      596 GTATGATCTCATTTAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGA 655
DB      148061 CTTTGAATGATCTCTGCTGACGACGACGACGACGACGATCTCTCCCGAGTTCGCG 148120
QY      656 GCTTTAAGAGATCCGAGACCTCTCTCCCTGCTGAGAGTCTGATACCGGCTTATG 715
DB      148121 GGAGTACGCGGCGACCG---CCTTACCGACCGCGCGGATCATCTTCCGCGCTACG 148177
QY      716 TACTCGCGCTGAGTCTGCGAAGAGTGAAGGCGCATCAACGATGATGATGATGATG 775
DB      148178 TCTTGGCCCGCGCGGCGCTTGTCCGACACGACGCGCACCGCGCGCTACTGGA 148237
QY      776 AGCGGCGATGCGCGAGCTTCCACCTCC----- 803
DB      148238 CGGAACGCGTGAACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148297
QY      804 -----GCCGACGCTTCCATGAGGCGCATCTACCTCTGAAGAGATCCG----- 851
DB      148298 CCGCGGTGACGACCCGCGGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148357
QY      852 -CTTCCGACACGAGGCAATGCTGCGTGGAGCTCTCTGGGAGTGAAGCGCGGTG 910
DB      148358 CGTTCCGCGCGGTGAGAGGTGCTGCTCCCGCGCGACCGGACCGGCTCACCGGACGTG 148417
QY      911 TCGGGAGCGCGGAGTGAACCCGACGCGGCGTATCTTGGCTGCAATTTTCCGAGTGTG 970
DB      148418 CCGCGCGCGCGGCGGTGACCCCGTGAAGGCGCTCTCAACCGCTTACCGCGAGAGGTG 148477
QY      971 GCGCGTGAAGCGCGAGCCCGCGCTTATGCGTGAATAGCGCTTCAACCGGCTCCCG 1030
DB      148478 GCGCTGTGTGCGCTCAAGCGGCTTCAACCTTAAAGTGTGCAACCGTCAACCGCGCG 148537
QY      1031 TCCATTCGCGCGTGAAGCATATCAACCGGAGCTTCAACGATGATGATGATGATG 1090
DB      148538 TGCACAGAGACATGACCGGCTCTGCTGCGGATTTACCTCGTGAACCTCTCGCGCGT 148597
QY      1091 ACACCACTCGCGACAGAGCTTGAACAGCGCGCTTAAGCGTATTCAGAGACGCTGTGG 1150
DB      148598 ACCTGACACACCGCGACCTTTCGCGGAAAGGACCGGCGCGCTGAGCGGACAGCTCTG 148657
QY      1151 AAGCATGATATCTGAGACGTAAGCGATATGATGATGATGATGATGATGATGATG 1210
DB      148658 ATGACCTGCGCCACCGCGCTTACCGGCTGGAAGTGTG-----TGGCGAATCTCTCC 148711
QY      1211 TGGGATTCGAAGAGCGGATGTTTCCCGTGTGCTCAAGAGCGCGCTTAAACAGCAG 1270
```

Db	148712	GCAGGGCCGGGCGCTCCGGGCTCATGCCGGTGGTCTTACACAGGGCCCTG-----G	148762
Oy	1271	TCGTTGGTGTACCTTCGTTGCAGAGGCTTGAACTCCGGTTACACACAGCGACATC	1330
Db	148763	GGCGGGGTCACCTCCGAGGCGGTAACCCCGGAGGTGAGTAAAGCGTACCCGAGCAC	148822
Oy	1331	CTCAGCTCTGGTGGATCATCAGCTCTACGAGACGATGGGGGCGTGGTCTGGCTGGG	1390
Db	148823	CGCAGGTGGGCTTCAGCTGCCAGGTATATCAACGCGGCGACACTTCAGCTTCTGGG	148882
Oy	1391	ACATGTCGACGGAGTGTTCGCCCGGACCTTCTGAGCAGCATCTCGAAGCTACGTG	1450
Db	148883	ACATCGGAGAGGGCGGCTCGCCACAGGAGCGGCGGACCGCATGTTCGAAGCGTACCG	148942
Oy	1451	TTTTTCTCCGGGGCTCACT-----GAGGAACCATGGGGTGAAACAG	1492
Db	148943	CACCTGTCGTTCCGTGTGCGCGGAGGCGAGGCGAGGCGAGAAAGGCTGGAGCGACCG	149002
Oy	1493	TGCGCTGTTCGTTCTCCGCTGCGCCAGCTGAAAGCGCGGCGGAGCGCAACG	1552
Db	149003	TCCGATCCCGCTGCGCGGCGACAGGGGCAAGCGGGCGCGGTCAACGACCGAGG	149062
Oy	1553	CGCTCTGAGCGAGCATACGCTGCACGCGCTGTTCGGGCGGGGTGAGCAGCTCCCA	1612
Db	149063	GCCCCCTCCCGGACCCCTGCTTCCAGAACCGGTACTCCCGGGGCGAGCACACCGG	149122
Oy	1613	TGCAGCTCGCGGTGGTGTGGCGCGCAACGCTACGTAAGAAAGCTTTCGCGCGTT	1672
Db	149123	ACGCGATGCGGTCCGCGACCCCGAACTGGCCCTCTCTACCGGCGAGCTGTACCCCG	149182
Oy	1673	CGCGCGACTTGGCGCGCGGCTGCGGAGCGAGGGGCAAGCCCGAACACATTTGGCGG	1732
Db	149183	CCACGAGACTCGCCGACACTGACGGGCTCCGGATGCGGCCCGGGGAAACGGTGCCA	149242
Oy	1733	TGGTATGGAAGAAAGGTTGGGAGCAGGTGTGGCGTTCTCGGGGTGCTGATCAGGGG	1792
Db	149243	TCGTGATGACAAAGGCTGGGGAACAGGTGTGCGCTCTTCCGACCTTCATGGCGGGG	149302
Oy	1793	CGGCTTACGTGGCGGATTCGATGCGGACCTACCGGAGGATATCCACTACTCTCGATC	1852
Db	149303	GCGCTTACTGCTGCGGTCCACAGCGGCCACCGCGGCGCCGCGGAGACAGATCATGGGG	149362
Oy	1853	ATGCTGAGGTAAAGCTCGTGTGTCGACGCGACCAATGCTGATGGCAACTGTATGCGCG	1912
Db	149363	ACGCGGGGCTCGGCAACCTGCTTCAACCAAGTCTGGCTCGCGGAATCGAAGACTCCCT	149422
Oy	1913	CGGGGATCCAGCGGCGGCTGCTGAGGAGGAGGCGGCGGTCGAAAGCGAGCAGCGTC	1972
Db	149423	CGACGTGTGCGCGGTGCGCGT-----GAACTCGTGGGAGGCGCACCGCGGACTTC	149476
Oy	1973	CGATGATGCCCAATTGAGACACTTCGGATCTCGCGTATGTATCTACCTCGGGATCCA	2032
Db	149477	CGCCCGCGCGCGCGGAGCCTCGAGCACTCGGCTAACGTCACTTACCTCGGGCTCCA	149536
Oy	2033	CAGGTTGCCCAAGGGGGTGATGATATATCGGGGTGCGTCAACACTATCTGAGCA	2092
Db	149537	CGGCGACGCCAAAGCGGTGATGATCAGCACCGGGCGCGCTTCACACCGTGAAGACA	149596
Oy	2093	TCAAAGACGCTTTCGAAATAGGGGCGGAGACAGGGTGTCTGGGCTCTCTCGCTAGCT	2152
Db	149597	TCAACCGCGCTTTCGCGGTGCAAGAGCGGACCGCTCTCGGACATCGCCGACTCGGCT	149656
Oy	2153	TCGATCTCTCGGTCTATGATGTGTTTGGGATCTCGCGCGCGGGCGGATCGATGTGGTGC	2212
Db	149657	TCGACTGTGCTCTTACGACTGTGTTGAGACCGTCCCGTGGGCGCACCTGTCTGTG	149716
Oy	2213	CGGACGCGTCCAAAGCTGCGCGGATTCGGGCGGATTTGGGCGAGTGTGATCGAACAGAAAG	2272
Db	149717	CCGGGTCCGACTGCGCGGCGACCTTCCCATGGGCGCAACTGTGTCCGGACTTGGGG	149776
Oy	2273	TCACGCTGTGAACCTCGGTGCGCGCGCTATGCGGATGCTGTGTGACATTTTGAAGGTC	2332

Db	149777	TCACGGGTGTGGAACTCCGTAACCGGGCAGACTCCACATGCTGTGGCACTGGCTCCG	149836
Qy	2333	GCCCCGATTCCGCTCGCTAGGTCCTCTCGGCGCTTTCCGCTGCTAGCGGGCAGTCCGATCCCGG	2392
Db	149837	AGCCCCCCACCGACGACGAGCTCTCCCTCGCGCTGGGCACTGATATCGGGCGAGCTGGATACCGG	149896
Qy	2393	TGGGCGCTGCTGGAGAGCTCCAGGCGCATCAGGCGCGGCGCTGTCTGGGTATCAGCGCTGGCG	2452
Db	149897	TTCGACTCTCCCGACCAAGGCCCGCTGAACTCTTCGCCGAACTGGAGATCGTCTCCCTGGCG	149956
Qy	2453	GAGCCACCGAAGCGTCGATCTTGCTCCATCGGCTACCCCGTAGAGAACTGTGACTATCGT	2512
Db	149957	GAGCCACCGAGGAGCTCCATCTGTGCTCATCGCCACCGCATCGGCGAGGTCTGACACCGCCC	150014
Qy	2513	GAGCGAGGATCCCTCCTACGGGCGCGTCCGCTGGCGCAACGAGCGTTTCCAGCGTCTGATGAG	2572
Db	150017	GCCCCAGATCCCGTACGAGCAAGCCCTCTGACCAACAGACGTTCCGCGTCTGACCGGC	150078
Qy	2573	CGCTCGAACCAGCCCGCTCTGCTGTGGTTCCGGGAGCACTTACATTTGGCGGGGCTCGGCTGG	2632
Db	150077	ACCTGCGGCGCGGGCGCCGAGTGGGTGGCGGGGAGAGCTGATACATCGGCGGGTGGTTCG	150136
Qy	2633	CACGTGGGCTACTGAGCGCGCATGAAAGAAAGACCGCGAAGAGCTTCTCTGTGCAACCCGAGA	2692
Db	150137	CCCTGGGCTACTCTCGGCGACGCGGAGCGCACGACACAGGTTCTTCACCGACTTGCCCA	150196
Qy	2693	CCGGGAGACGCGCTCTACAAAGACCGGCGATCTTGAGCGCTACCTGCGCCGATGGAAACATCG	2752
Db	150197	CCGAGAGACGCGCTCTACCGGCAACGGGCGACCTGGCGCTACTCTGCCCGAAGGGGACATCG	150256
Qy	2753	AGTTCATGGGCGGTGAGAGCAACCAATCAGCTTTCGGGATACCGCGTTAGCTGGGG	2812
Db	150257	AGTTCCTGGGCGGGAGGACCGCCGAGATCAAGATCCGCGAATACCGGGTCTGAACTGGCG	150314
Qy	2813	AAATCGAGAAACCGTCAAGTGCATCCGAAAGTACGAGCGCGGATTTGGTCCCGTCG	2872
Db	150317	AGGTGAGAGCGCGCTTACAGATCCACCCCGCGGTGCGCGGCGCGGTCTGTGTGAGCG	150378
Qy	2873	GGAACGACCGCGCGCAACAAGCTCTTCTTAGCTATGTGTCCCGGAGGGCACACGAGAC	2932
Db	150377	ACTCGGGCGCGGGCGGGCGGCGGCGCTGTGGCGCGTGTTCGAGAGCGCACGCAAGAGACGGCG	150436
Qy	2933	GCGCTGCGGAGCAGGACCGCGAGCTCTCAAGACCGAGCGGATTCAGCGGAGAGCACACGGCG	2992
Db	150437	GAGCCCGCGGTGACGAGCGCGCACAGCGCGGACGCGACGCGCAAGCGGTACGCGAAG	150496
Qy	2993	CCGAAAGCGACGCGC 3006	
Db	150497	CGGGGCGGGCGCTC 150510	
RESULT 15			
AL646066			
LOCUS	Ralstonia solanacearum	189050 bp	DNA linear BCT 11-APR-2003
DEFINITION	segment 10/19,		
ACCESSION	AL646066	AL646066	
VERSION	AL646066.1	GI:17428714	
KEYWORDS			
SOURCE	Ralstonia solanacearum		
ORGANISM	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;		
	Burkholderiaceae; Ralstonia.		
REFERENCE	1		
AUTHORS	Salaouabati, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S.,		
	Arlat, M., Billault, A., Broclet, P., Camus, J.-C., Cactolico, L.,		
	Chandler, M., Choisme, N., Claudel-Renard, C., Cunha, S., Demange, N.,		
	Gaspin, C., Lavey, M., Moisan, A., Robert, C., Saurin, W., Schlex, T.,		
	Siguiet, P., Thebaud, P., Whalen, M., Wincker, P., Levy, M.,		
	Weissenbach, J., and Boucher, C.A.		
TITLE	Genome sequence of the plant pathogen Ralstonia solanacearum		
JOURNAL	Nature 415 (6871), 497-502 (2002)		
MEDLINE	216181879		

PUBMED 11823852
REFERENCE 2 (bases 1 to 189050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, UMG CNRS
118 Route de Nardonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
location/Qualifiers
1. 189050
FEATURES
source /organism="Ralstonia solanacearum"
/mol_type="genomic DNA"
/strain="GMI1000"
/db_xref="taxon:305"
170. .3325
/gene="RSC1698"
/note="synonym: RS02885"
170. .3325
/gene="RSC1698"
/function="Elements of external origin; phage-related
functions and prophages"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE BACTERIOPHAGE PROTEIN"
/protein_id="CAD15400.1"
/db_xref="GI:17428715"
/db_xref="SPTREMBL:O8XYR0"
/translation="MRNIIGYGGKGGSSSTPTPEADPSLHSVAVARVLVSEGEIA
GLVNGLSIYLGTPLANADGTINFONVADVDPGSDOEALRGPPSVSESAVVEL
TATMPWVAINTOISAVRVOISVPLSGSDPSGNNINGRVVVAIDILAVGALGOV
LATADGKTSKTRTRHRELPRAKTGTIRRTTPNANSGTIADVTIRIESIAVID
AALRYPNLSALIGRIDARQSSVPTRSYHLRGRITRPENYDPATRTYGLMDGTFKV
ASNNPAMWFYDIVLHKRGGLDRVASADVDWMALYQIQYCEDELVPDGRGQBERFT
NCYCIQORNDAPVLQDLASVFRGMAYMSNVAVADMPASAYLFPHGNVDERFT
YAGSARARAKTVALVMDPGQYRAKVEHVDGGIARYSIQTEVTAFCGTSSOA
ORVGOMILLTSRLTETVSPKYGDAVVMPGAIIENVADPARAGTNGRVRVSADGRT
VYVDRPVYVRBEGDTILINPDGTAORRTISRDGSLTYSADMSQPVQAEAVMSESA
DKITQLFRVSVTESDGLTFDISALEYNOSKPSAIDNGTRIEARPEVSLPESVQAFPT
DYTLATYSVVDGIAITMMVAMQQAASAVALTEVRNDGEMVTAIPGSGSLERN
IYAGTYVAVRAINALDVMSLPAYSAEFLTKTSPPVVGTLLIATAIFGIRLMAF
PTGPDVERTELMSKAPRASATKLGDFAPANTHTLMLGAAATFEYMAFLVDSKQ
NVGAVYPSGNGVIGMSDRAPIILSYMGOICGLDGLLSAVDSIDPMAKSGEDF
AGDGVYAGIVTSQVLEBANRAIAQOTTMTATYAQSSAAVQAQAAVABEGGILAA
MTTIQTIAANGRIYIAGIFGVENDKGVTEQVLIADRFVAILPHNGNVNVTPTVVQ
GGQVPMDSAFIDGTITSKIDGSIQSSNFVAGQGTGWLKSGLFENNGNGAGRRVD
TSLATQIYDANGTLRLGVW"
3662. .4318
gene /gene="RSC1699"
/note="synonym: RS02888"
3662. .4318
CDS /gene="RSC1699"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11

gene /product="HYPOTHETICAL PROTEIN"
/protein_id="CAD15401.1"
/db_xref="GI:17428716"
/db_xref="SPTREMBL:O8XYO9"
/translation="MPAGLTYYGHHGFVQITAEYGNLALRQKSTVTPDGGSGRGSITFS
ASRPFCVQSANPVALLSGNSGADMTVFPAPQAPPTVVPFEPAPSGCPGLQV
RADGLIARDSGLAIVKSGVVTTPSGAPAYQAMASPLTGNTVAAACSYRTGTGYAV
PGDTLTPADHVITYTANAGLKLQYTRGAMNGDGSGLKIDTANPQIYVLVDVSQL
"
4387. .4878
gene /gene="RSC1700"
/note="synonym: RS02889"
4387. .4878
CDS /gene="RSC1700"
/EC_number="3.2.1.-"
/function="Elements of external origin; phage-related
functions and prophages"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE BACTERIOPHAGE ENDOLYSIN PROTEIN"
/protein_id="CAD15402.1"
/db_xref="GI:17428717"
/db_xref="GOA:O8XYO8"
/db_xref="SPTREMBL:O8XYO8"
/translation="MPYTDPSVLGGRNLAFLDMVGFSEGTGNGRQPTRDGVDLVG
GLFVSYADHPRILVELPRIGKSTAGRYOLLARYDPPYRROLRVLDFGPAADAI
VQOIRERCALDIDQAGRLAAIACKNIMASLPBAGYQGHQKEBALRAHLLCGCGE
VKG"
4878. .5186
gene /gene="RSC1701"
/note="synonym: RS02890"
4878. .5186
CDS /gene="RSC1701"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE LIPOPROTEIN TRANSMEMBRANE"
/protein_id="CAD15403.1"
/db_xref="GI:17428718"
/db_xref="SPTREMBL:O8XYO7"
/translation="WTSEBNVAAKVGSGSALGSVIALRFPGTWQRMLSFVSSLGI
GCLAGGAVERFLLVPSYTHMLAVASAVFGLAIVNNAMQOIPILTLDRRFLAKE
"
5189. .5464
gene /gene="RSC1702"
/note="synonym: RS02891"
5189. .5464
CDS /gene="RSC1702"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD15404.1"
/db_xref="GI:17428719"
/db_xref="SPTREMBL:O8XYO6"
/translation="MLTISINQVANAIIFVGSLSMAVLTHQVPTRTGALVLAVNFAA
LGNIVAHGTCHSWPEVALNVAVLVCVMAFRLREVRFRFSALGKGRG"
LGNIVAHGTCHSWPEVALNVAVLVCVMAFRLREVRFRFSALGKGRG"

```
gene      5464..6009
          /gene="RSC1703"
          /note="synonym: RS02892"
          5464..6009
          /gene="RSC1703"
          /EC_number="3.4.-.-"
          /function="elements of external origin; phage-related
          functions and prophages"
          /note="Product confidence: probable
          Gene name confidence: hypothetical
          predicted by Framed"
          /codon_start=1
          /evidence-not_experimental
          /transl_table=1
          /product="PROBABLE BACTERIOPHAGE-RELATED SIGNAL PEPTIDE
          PROTEIN"
          /protein_id="CAD15405.1"
          /db_xref="GI:17428720"
          /db_xref="GOA:Q8XY05"
          /db_xref="SPTREMBL:Q8XY05"
          /translation="MPGIGKTAIAAATIMLGLVGAAGATWQAHYKDIAFLORD
          QAERDROARADENQJREKREKAGLEQLDARHRELNKTERAAALAAAG
          TRQLSIRARCTAPADGVPSITSGAGLDGGATAILYGEDADIQALVADADIAAL
          RFGGRRAIQTRDTSVDAR"
          complement(6117..7304)
          /note="predicted by Homology"
          /insertion_seq="ISR018 or RS05932"
          complement(6273..7238)
          /gene="tISR018"
          /function="elements of external origin; transposon-related

repeat_region
gene
CDS
Query Match      14.0%; Score 593.6; DB 1; Length 189050;
Best Local Similarity 52.6%; Pred. No. 4.3e-84;
Matches 1510; Conservative 0; Mismatches 1309; Indels 51; Gaps 8;

OY      206  ACGCTCGGTTCTCTCTCAACAGATCCAAAGATCTTACTGCTGGCCGGACAGAGAGCT 265
Db      159271  ACGACACTTGCCTGACCCACATCCAGACGCTTACTGATCGGCGCGCCAGGCC 159330
OY      266  TTACGGTCCCGAGCGGG--ATCCAGGCTTATCCGGAATAGACTGTAGAGATCGACG 322
Db      159331  TGGGGCTGGGGGGTGTCAATGCCACGTCTTACTGATGGGTCTCGCCAAACCTGGACC 159390
OY      323  TGGCAGGCTGAGCGCGGCTTTTCGAAAGTGTGCGCGGCAAGCATGCTTCGGGCC 382
Db      159391  TGCACGCTTCGACGAGGCTGGAACCGGATCGTGGCGGCAAGCATGATGCGCGCG 159450
OY      383  ACAGGCTGCCGACATGATGAGAGATCGAGCCCTAAAGT---GAGCCGCAATCGAGA 439
Db      159451  TGTGACGCTCGGATGAGGCGAGGCAATCTTCGACCGCTCCGCACTACGCCATCGGA 159510
OY      440  TCATCGATCTGCGGGGCTCGACCGGACACAGGGAAGAGGCTCGTGTGTTGCGAG 499
Db      159511  CCGAAGACTCTGCGCGGACGCGCCGATCGCGTGGCGGCGAGCTCCGCAAGAGCGCG 159570
OY      500  ATGCGATTCGACCGCATCTATGACACCGAGCGCCCTCGCTCTATCAGCTCGCGCG 559
Db      159571  CCGGCACTGCGGGGAGGTGTCAACGCGGAGCGGTGGCGCTGTTCGACCTGGCGGCA 159630
OY      560  TTGGCGCTGAGAGAGCGGAAACCGCTCTGCTGCTCAGTATGATGATCTCTTAACG 619
Db      159631  CGCGAGCGCGCGCGCGCTGACCGGCTGCACTTCGATCTGCTGTGATGTTTCGACG 159690
OY      620  TAGCAGGCTGCTCATCTCTTCAAGAGACTGAGCTTCAAGCAAGATCCGAGACT 679
Db      159691  TGCAGAGCTTCCATCTCTGCTGCGCGGAGCTCGAGCGAGCTACCGGACCGCGCGCG 159750
OY      680  CTCTCCCTGCTCTGAGGCTCTGTACCGCGATATGTACTCGCGCTGAGGCTTCGCA 739
```

```
Db      159751  AGCTGCCCGCCATCGGCTCAAGTTTCGGGAGCTACCTGCTGCGCGCGGAGCGCGCG 159810
OY      740  AGTCTGAGGCGCATCAACGATCGATGATTACTGGAAGCGGCGATCGCGAGCTCCAC 799
Db      159811  GCGAAGCAGCTTATCTGGCCGACAGAGGCTGTGTGGAGCGCCGCTCTGCAAGATCGCG 159870
OY      800  CTTCGCGGACGCTTCGATGAGAGCGGATTCATCTACCTGGAAGAGATCCGCTTCCGC 859
Db      159871  CGCCCCGCGGCTCCGCTCGCGCGGTGGGCTCGAGCTGTGGCGCGGCGCATTCGCGCG 159930
OY      860  ACAGGAGCAATGCTGCGCTCGGACTCTCGGGGTCGATTGAACGCGGTCTCGGGAGC 919
Db      159931  GCTTCAGACATCGCTTGACCCACAGGCGTGGCAGCGGCTGTGCTTTCGCCCGAGG 159990
OY      920  GCGGCTGAGCCCGACGCGGCTCATCTGCGCTGATTTTCGAGATGATCGGGCGCTGA 979
Db      159991  CCGGGCTGACCGGCTCTGCTGCTGCTGCTTCAACCGGCTTCGGAAGTGTCTCGCGGCTGG 160050
OY      980  GCGGAGCGCCCGGTTTACGCTCAATPAACGCTTTCAACCGGCTCCCGCTCATCGCG 1039
Db      160051  CCGGCGACGCGCGCTTACGCTCAACCTCAACGCACTTCAACCGCGCGCGCTGATCGCG 160110
OY      1040  GCGTGAACGATATACCGGGGACTTCACGCTGATGCTCTCTGGAATGCAACCACTC 1099
Db      160111  ATGTGAGCGGCTGTGCGGATTTCACTCGGTGCTGCTTCAAGGATCGATGCAACG 160170
OY      1100  GCGACAGAGCTTGGAACAGCGGCTAAGGCTATTCAAGAGCAGTGTGGAGAGGATGG 1159
Db      160171  CGCCGCTGCGCTTGGCCAGGCTGCGCGCGCACCCAGGCCCACTTGTGGCCAGCTGT 160230
OY      1160  ATCACTGCAAGTAAGCGGATATCGAGGTCCAGCGAGGCGGCGC-----1205
Db      160231  CGCACAGCGGCTTGGGGGGGTGAAGCTCGGCGGAATGCGCGGACCGCGCGGAG 160290
OY      1206  -GTTCTGGGATCAACAGAGGCGCATTTTCCCGTGTGCTCAACGAGCGCTTAAC 1264
Db      160291  ACGGCGCGGATGAGCGGCGGCGGCTGATGCGGCTGTGCTTACAGAGCTCTTGAGCA 160350
OY      1265  AGCAAGTGC---TTGATGTCACCTCGTTGAGAGGCTCGGAATCCGGTGTACACGACA 1321
Db      160351  TGAATCTGACCTCTGATCTGCTGCGGCGGACCTGCTGAGGACCGGAATTCCTTACA 160410
OY      1322  CGCAGACTCTCACTGCTGCTGATCATGCTCTACGACAGATGAGGAGCTCTGCC 1381
Db      160411  CTTGACGCGCCAGACTGTGGTTCGACACAGAGATGATGTCCGAAAGGACAGCTGGAA 160470
OY      1382  TCGGCTGGAACATGTGAGCGAAGTGTCCCGCCGACCTTCTGAGAGCATGCTCGAAG 1441
Db      160471  TCAACTGATCTGATGAGAGCACTGTTCGCGAGGCGATGTGATGCTGATGTTTCGAGG 160530
OY      1442  CGTACGTGTTTTTCTCCGCGGCTCACTGAGG-----ACATGGGGTGAACAGGTGC 1495
Db      160531  CTATTTGCGGCGGCTGACCATCTGAGCCAGAGATGCGCGCATTTGCAAGCAGATGC 160590
OY      1496  GCTGTTGCTTCCGCTGCGCAGCTAAGAGCGCGGCGAGCGCAACGCAACGCGC 1555
Db      160591  CTGTGAATTCGCGCGAGCGAGCGCGGATGCGCGAGGCGTGAACCGGACCGCGCTAC 160650
OY      1556  TGTGAGCAGATACGCTGACGCGCTGTTCGCGCGCGGCTGACAGACTGCCCATGC 1615
Db      160651  CGCTGCGGAGTCCGCTGACCGGAGCTTCTTCGCGGAGGCGGATGCGGACGCGGACG 160710
OY      1616  AGCTCGCGGTGTGTGCGCGCGCAAGCGCTCAAGTACGAAGACTTTTCGCGCGTTCCG 1675
Db      160711  CGTGGCTTGCGCACCGCCATGCGAGAGACTATGCGGACTGAGCGCCCGCGCGCG 160770
OY      1676  GCGACTTTGGCGCGCGCTGCGCAGAGAGGAGGCGACCGCGCAACATTTGGTCCGCGTGG 1735
Db      160771  TGGGCGAAGTGGCGAGCTTGGCGAGGCGGCGGATGATGCGCGGATCTGTTGCTGGCG 160830
OY      1736  TGATGAGAAAGAGCTGGAGAGAGTGTTCGCGGTTCTTCGCGGATCTGAGTCAAGCGCGG 1795
Db      160831  CATTGCCAAGGAGATTCCGCGCTCGCGCGGCAATCGGATGTCGCGCGCGGCGCGG 160890
```

QY	1796	CCCTAGTGGCGATTCGATGCGACCTTACCGGCGGAGGATTCGACCTACCTCTCGATCATG	1855
Db	160891	TGCTGTGCTGCTTTGGCCAGCAGACATGCCCCGGCGCGCATGCAAGCATTTCTGATGGCG	160950
QY	1856	GTGAGGTAAAGCTCTGCTGTGACGCGACGACCTATGCTGAGTGGCAAACTGTCAATGCGCCCGG	1915
Db	160951	CAGGTGTCAAAGCCGCGCTTTATATCGCGCAGCGGAGCTTCAACGCGCGGATGCGCAGCC	161010
QY	1916	GGATTCAGCGGCTGTCTC-----GTGAGGAGGCGCGGCGTTCGAAAGCGACAGGACACAC	1969
Db	161011	CCGTGCGGCTGGCTGTGCTGTGCGCGGAATTCGACGCGGCGCGGTGTACGGAAGCGCGCGCTGC	161070
QY	1970	CTCCGATGATGCGCAATTCAGACACCTTCG-----GATCTCGGTAATGTATCT	2017
Db	161071	AGCGGGAAATTCGACCGCGCACGTCGCGTCCGCCCGCTCGATGCGCTGCGCTACATATCT	161130
QY	2018	ACACTCTGGGATTCACAGGGTTGCCCAAGGGGTGATGATTCGATCATCGGGGTGCCGTCA	2077
Db	161131	ATACCTCGGGCTTCGACCGGCACGCGCCCAAGGGGTGGCGATTCACGCGCGCGCGCGCA	161190
QY	2078	ACACCATCTCGAACATTCACAGAGCGGCTTGAATATAGGGCCCGGAGCAGGGTCTGAGCCG	2137
Db	161191	ACACGCTGGAACCTGTTCACGCGCTGCTGAGCGCTGACGCGCGGACGACCGGCTTGCGCA	161250
QY	2138	TCTCTCTGCTGAGCTTTCGATCTCTCGGCTATGATGTGTTGGGATCTTGCGCGCGGGCG	2197
Db	161251	TCTCGGCGCTGGGCTTCGACCTTGCGACCTCATCAGATGTTCTTCGCGATGCTGTGCGCGGGCG	161310
QY	2198	GTACGATCTGTGCTGCGCGGACGCGGTCCAACTCGCGCGATCCGCGCGCTTGTGGGACGATTTGA	2257
Db	161311	GCGCGGTGATCTCTCCCGACGAGCGGGGCAATGGCGGATCTGTGCACTGGCTGACCTGG	161370
QY	2258	TGGAACGAGAGAGGTGACGAGTGTGGAATCTGGTGTGCGGCGCTGATGCGGATGCTGTG	2317
Db	161371	CGCACCGGATACGGCGGTCAAGCTGTGGAATCTGGTGTGCGGCGCTGTGCGGAGATGCTGTG	161430
QY	2318	AGCATTTTGAAGGTGCGCCCCGATTCGCTGTGAGTGTCTGTGCGGCTTTGTGCTGTGAGCG	2377
Db	161431	AGCGGCGCGGCATTCAGCGGTGCGGGGCTCTCGGCATCTGTTGCTCGGTGATGCTCAGCG	161490
QY	2378	GCGACGTGATCCCGGTGGGCGCTGCGCTGGGAGGTCCAGGCGCATCAGGCCCGGCGTCCG	2437
Db	161491	GCGACTGATTTGGCTGTGACCTTGCACGCGCTGTGCGGCACTGGCGCGCACACCGCC	161550
QY	2438	TGATCAGCTTGGCGCGGGCGCACCGAAGCGTGTGATCTGTTCATCGGGTACCCCGTAGGA	2497
Db	161551	TGGCGGCTCTCGGCGGGCGCACCGAAGCGCGGATCTGTGTGAATCTGGTTGAGGTGCGGG	161610
QY	2498	ACGTGACCTATCTGTGGGCGAGCATCCCTTACGCGCGTTCGCTGCGGACACGACCTTCC	2557
Db	161611	GGGTGTGATCCCGCTGGCGCTCTCAATACGATCGGGTTTCCGTGTGCAACCACTCTTACC	161670
QY	2558	ACGTGCTCGATGAGGCGCTCGAACCGCGCGGCTCTGGGTTCCGAGGCACTCTACATTG	2617
Db	161671	GGGTGTGTGATGATCCACTGTGCGGATTTGCCCGGACTGGGTGTGATGCGATTGTGCATATG	161730
QY	2618	GCGGGGTCTGGGCTGGCACTGGGCTACTGGCGGATGTAAGAGAAAGCGCGACGACTTCC	2677
Db	161731	GCGGGCGCTGGCTGTGCTTGTGTCTACTGTGCGCATCCGACGCGCACCGCGGAGGCTTTCA	161790
QY	2678	TGCTGCACTCCGAGACCGGAGGAGCGCTCTACAAAGACCGGCGATCTGGGCGGCTTACTCG	2737
Db	161791	TACCCACCCCGCGACGCGAGAGCGCTGTATCCGACCGGCGACCGCGCGGCTTACTGGC	161850
QY	2738	CCGATGAAACATTCGAGTTTCAATGGGGCTGTGAGGACAAACCAATCAAGCTTTCGCGATACC	2797
Db	161851	CGAGCGGCTGTGATTCGAAATCTCTCGGGCGGCTGTGACAGCCAGGTCAGGCTTCGCGGACACC	161910
QY	2798	GCGTTGAGGCTCGGGGAAATTCGAGGAAACGCTCAAGTTCGATCCGAACTGACGAGACGGCG	2857
Db	161911	GCATCGAGCTGGCGGATGTGAAGCCGCGCTGTGCGAGCACAGCGGATTCAGAGCGCGCG	161970

```

Qy      2858  CGATTGTCGCCGTCCGGAAACGACGGGCGACAAACCTCTTGTAGCTCATGTGAGTCCGG  2917
Db      161971  CCGCCACGGTCTGCCCCACACGAGGGGGCGGGGACACAGTGGCCCGCTTGTGTGTCGG  162030
Qy      2918  AGGGCACAACGAGACGCGCTGCACGACAGACGCGACGCTTCAAGACCGAGCGGATGACG  2977
Db      162031  GACACGAGCCCG---GCTTCTGCGCGACGGTCCGCCACACATGACCGCTGGCAGGCCCTGG  162087
Qy      2978  CGAGAGCACACGCGCCGCCGAAGCGGACGCGCTTAGGAGACGCGGACGAGGGGTG  3027
Db      162088  CGGCGGCGCTGCGCGTCCGGGGGGGGCGCGCTCCCGCGCGGCGCAAGGTG  162137

```

Search completed: October 4, 2003, 03:30:57
Job time : 15178.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:01 / Search time 2488.11 Seconds
(without alignments)
12307.996 Million cell updates/sec

Title: US-10-014-717-1_COPY_62369_63628

Perfect score: 1260

Sequence: 1 atgacacagagacagcagca.....agccctccaaagcgcgacag 1260

Scoring table: IDENTITY NUC

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rtd:*
26: em_ges_phg:*
27: em_ges_vrl:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	9.1	1285	29	B2567880 pac62-164
2	84.6	6.7	29	B2675691	B2675691 PUBDH78TD
3	63	5.0	330	28	AY174775 AY174775
4	55.4	4.4	1127	29	B2548452 pac61-60

5	54.8	4.3	925	29	CNS0091P	AL053013 Drosophila
6	54.2	4.1	1258	29	B2573154	B2573154 msh2_2968
7	52	4.3	609	12	BI716406	BI716406 1031010A1
8	51.8	4.1	925	29	CNS0091P	AL053013 Drosophila
9	50.8	4.0	425	10	BG488189	BG488189 RHIZ2_60
10	50.6	4.0	1170	29	B2569411	B2569411 pac62-164
11	50.2	4.0	512	14	CD045174	CD045174 pshB018x
12	50.2	4.0	512	14	CD045351	CD045351 pshB019x
13	50.2	4.0	596	14	CD260296	CD260296 pshB008x
14	50.2	4.0	610	14	CD043612	CD043612 pshB009x
15	50.2	4.0	623	14	CD262676	CD262676 pshA018x
16	50.2	4.0	715	14	CD041124	CD041124 pshB001x
17	49.8	4.0	1242	29	B2579496	B2579496 msh2_6381
18	49.6	3.9	638	14	CD460072	CD460072 CRUS5_4D0
19	49.6	3.9	691	10	BE559050	BE559050 HV CEP002
20	49.4	3.9	932	29	CNS00720	AL066742 Drosophila
21	49.2	3.9	402	10	BG559248	BG559248 RHIZ2_52
22	49.2	3.9	665	12	BM370205	BM370205 ERT08_50
23	48.8	3.9	788	14	CB668866	CB668866 OSJNE0160
24	48.4	3.8	899	13	BX360973	BX360973 BX360973
25	48	3.8	641	9	AU092915	AU092915 AU092915
26	47.8	3.8	572	14	CA404475	CA404475 E01N0518
27	47.8	3.8	583	10	BE761796	BE761796 947006H01
28	47.8	3.8	799	29	B2977578	B2977578 PUGIS17TD
29	47.8	3.8	838	29	B2977576	B2977576 PUGIS17TB
30	47.2	3.7	389	10	BE229868	BE229868 99AS102 R
31	47.2	3.7	406	9	AU063730	AU063730 AU063730
32	47.2	3.7	413	9	AU093702	AU093702 AU093702
33	47.2	3.7	417	14	D22231	D22231 RICC10540A
34	47.2	3.7	421	9	AU069623	AU069623 AU069623
35	47.2	3.7	442	9	AU182095	AU182095 AU182095
36	47.2	3.7	442	9	AU182164	AU182164 AU182164
37	47.2	3.7	446	9	AU182068	AU182068 AU182068
38	47.2	3.7	449	9	AU161375	AU161375 AU161375
39	47.2	3.7	549	10	BF287873	BF287873 100067D0
40	47.2	3.7	643	9	AW585283	AW585283 707094C03
41	47.2	3.7	707	9	AU161432	AU161432 AU161432
42	47.2	3.7	719	12	BM600174	BM600174 17006870
43	47.2	3.7	738	14	CB653178	CB653178 OSIEB15E
44	47.2	3.7	741	14	CD428040	CD428040 ETH1_32 D
45	47.2	3.7	764	14	CB668867	CB668867 OSJNE0160

ALIGNMENTS

RESULT 1
B2567880
LOCUS
DEFINITION
pac62-164_7256.xl pac62-164 Pseudomonas aeruginosa genomic clone
pac62-164_7256, genomic survey sequence.

ACCESSION
B2567880
VERSION
B2567880.1 GI:27199936

KEYWORDS
GSS.

ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

REFERENCE
1 (bases 1 to 1285)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

AUTHORS
Burns, J.D., Kaul, R., and Olsen, M.V.

TITLE
Whole-Genome-Sequence Variation among Multiple Isolates of

JOURNAL
J. Bacteriol., (2002) In press

COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954

FEATURES
Email: 2066857244
Email: craymond@u.washington.edu
Class: Shotgun.

Location/Qualifiers

	source	1.. 1285 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="2-164" /db_xref="taxon:287" /clone="pac82-164_7256" /clone_1fb="pac82-164" /note="clinical isolate 2-164 whole genomic shotgun library."
BASE COUNT	219 a 405 c 355 g 289 t 17 others	
ORIGIN		
Query Match	9.1%; Score 115; DB 29; Length 1285;	
Best Local Similarity	53.3%; Pred. No. 1.6e-16;	
Matches 264; Conservative	0; Mismatches 230; Indels 1; Gaps 1;	
Oy	670 GACGCTTGACAGTCTGTCTTCAGGCCGAGCCGACCGCAGCATGAGCACCAAGGAG 729	
Dd	95 GACGCTACAGCGGGCTGTGTAGAGCCCCGACGACCGGCAGTTAGGAAAGCGGAA 154	
Oy	730 CTGTGTGCGCTCTGTGGGTGCATTTATTCCTGTGTGCACCAGATTACCAGATCTTACTTC 789	
Dd	155 CTCGTCTTCATAGGCCACCTGTGTATGTATGAAGCGCTTGTGAGCACATGACATGATC 214	
Oy	790 GCCTTCGCTGTGTCTAACCTGTGTGCGGTGCCCAGAGGCTGTGAGCTGTGAAGCCGAG 849	
Dd	215 GGCAACGCGCTGTGTACCCCTGTGTGTCAAACCGGAGCACTGTGCTTCTGCGGGCGAG 274	
Oy	850 CCGGGGCTCATAGAGGACGCGCTGATGAGGTCGCGCTTGCACATATCTCTGAGATA 909	
Dd	275 CGGAACTCTCTCCCAAGCCATGAGAATCTGTTCGCTCAGCAAGTCGCGTGCAGCC 334	
Oy	910 GGAACGTGCGTTTGGCAGGACGACCTGTGAGTACTCGGGGACATGCATCAAGAAAGG 969	
Dd	335 TCGATGTTTGCGCTTCAACGTGGAAAGCGTGAACTGACGAGGGGTCAACATTTCGCGCGC 394	
Oy	970 GAGATGCTTTTCTCTCTGATCCCGGAGCCCTGAGAGATGTGGACCTGTATTCTTCAGGGCA 1029	
Dd	395 GAATATCATCTGTGTTCCTCAACTGACCGGCAACAGATGCGAGCGCTTGCAGCATCC 454	
Oy	1030 GACGTTTGTGATGTGTCGACGAGGACGAGGCGGACCTGTGCGTACGTAGAGGCCCCCAT 1089	
Dd	455 GACGCGCTCGACCTGACCGCGAACAACGATGGCCATCTCGGCTTAGCGTTGGGGGTGAC 514	
Oy	1090 GTCTGCCCCGGGGGTCTCCCTTGCTGTGCTCTCGAAGGCG-GAGATGCGCGTGGGACCATCTT 1148	
Dd	515 TACTGCGGTGCGGGCTCTGCTGTGCGCGGTGAGGGGCGGATTCGCCATTCAGCGCTGCT 574	
Oy	1149 CCGTAGGTTCCCGCA 1163	
Dd	575 CCGCGGCTTCCCGA 589	
RESULT 2		
LOCUS	BZ675691	
DEFINITION	PUBDH78TD ZM 0.6 1.0 KB Zea mays genomic clone ZMBR7029ML2,	
ACCESSION	BZ675691	
VERSION	BZ675691.1 GI:28226790	
KEYWORDS	GSS.	
SOURCE	Zea mays	
ORGANISM	Zea mays	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 975)	
AUTHORS	Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utecherack,T., Resnick A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and Bennetzen,J.	
TITLE	Maise Genomics Consortium	
JOURNAL	Unpublished	
COMMENT	Contact: Cathy Whitelaw	
	TIGR	

FEATURES					
source					
	location/Qualifiers				
	1..975	/organism="Zea mays"			
	/mol_type="genomic DNA"				
	/strain="873"				
	/db_xref="taxon:4577"				
	/clone="ZMWBta029M12"				
	/clone_1lb="ZM 0.6-1.0 KB"				
	/note="Vector: PCR4-TOP0; Site 1: EcoRI; 0.6-1.0 kb high Cor selected genomic DNA library"				
BASE COUNT	173 a	323 c	302 g	177 t	
ORIGIN					
Query Match	6.7%;	Score 84.6;	DB 29;	Length 975;	
Best Local Similarity	49.3%;	Pred. No. 1.9e-09;			
Matches 250;	Conservative 0;	Mismatches 254;	Indels 3;	Gaps 1;	
Oy	668 ATAGACGCTTGAACATGCTGTCTTCAGGCCGACGGCCGACGCGCAGAGAGCTGAGACGCAAGG	727			
Db	216 ATGACTTTATACCGCCCTTACCACCCGACGAAAGAAGCGAGTCCCTTCAAGAGAGC	275			
Oy	728 AGCTGTCGCGCTTGTCGTCGATTATCGCTGTGCACCCGATACACGATCTACCTTA	787			
Db	276 AATTGCTGCTGTTGTGATTTGCGATCTCTGTCGCGGTTACGAAGCAGTGCTTCGACA	335			
Oy	788 TCGCGTTGCTGTCTCAACTGTGTGCGGTCCCGACGGCGCTCGAGCTGTGAAGCCG	847			
Db	336 TCCCCACTTATCCAGAGTCTGCTGTGACCAACCCCGCAGTGGCAACAGCTCAAAGCCG	395			
Oy	848 AGCCCGGAGCTCATGAGAACGCGCTCGATGAGGTGCTCGCTTCGACATATCCTTCAGAA	907			
Db	396 ATCCCGACGATACCGCAGCGGTCCAAAGCTGCTGCTTACATCCCGCTGGCTCGG	455			
Oy	.908 TAGAAA--CTGTGCGTTTTGCGCACGAGGACCTGAGTACTGCGGGCATTCATCAAGA	964			
Db	456 CGGGAATGTTTGTGCACTACGCGCTGGAAACATTCAGTTCGGTGAACCTTGGTGCGCC	515			
Oy	965 AAGGGAGATGCTTTTCTCTGATATCCGAGCGCCCTGAGATGGGACTGTATTTCCCA	1024			
Db	516 AGGCGAAGCGGTGTGCTTCGATTCGCGCGCGCAACATGACCCGCGCTTGAA	575			
Oy	1025 GGCCGACGCTGTGTGATGTGCGACGAGGACACGCGCGGACCTGCGCTACGTAAGAGCC	1084			
Db	576 ACCCAGAGGCGCTGACCTGCAACGAGATGCAAGCGCCCATTTTCGCTTTGGCCACGGCC	635			
Oy	1085 CCATGTCGCCCCGGGGGTGTCTTCTGCTCGCTTCGAGCGGAGATGCGCGTGGGACCA	1144			
Db	636 TGCACTCACTGATGGCTCAAGCCCTTGCGCGGGGTGATAATTGCAAGAGAGCGCTGACGCC	695			
Oy	1145 TCTTCGATGATCCCAGATGAAGC	1171			
Db	696 TGTGTGTGCGCTGCCAGCCTGAGC	722			
RESULT 3					
LOCUS	AY174775	330 bp	DNA	linear	SSS 13-JAN-2003
DEFINITION	AY174775 Mycobacterium avium subsp. paratuberculosis DNA				
	Mycobacterium avium subsp. paratuberculosis genomic clone Mptbel.15				
	' genomic survey sequence.				
ACCESSION	AY174775				
VERSION	AY174775.1	GI:27689314			
KEYWORDS	SSS.				
SOURCE	Mycobacterium avium subsp. paratuberculosis				
ORGANISM	Mycobacterium avium subsp. paratuberculosis; Actinomycetales; Bacteria; Actinobacteria; Actinobacteridae; Actinomycetaceae; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium				

REFERENCE avium complex (MAC).
 1 (bases 1 to 330)
 AUTHORS Nielsen, K.K. and Ahrens, P.
 TITLE Subtractive hybridization and bioinformatics: combining genetic
 methods to obtain new information on species-specific elements of
 Mycobacterium avium subspecies paratuberculosis

JOURNAL Unpublished
 COMMENT Contact: Nielsen KK
 Bacteriology
 Danish Veterinary Institute
 Bulowvej 27, Copenhagen V, 1790, Denmark
 Email: kni@veti.dk
 Class: unknown.

FEATURES
 source Location/Qualifiers
 1..330
 /organism="Mycobacterium avium subsp. paratuberculosis"
 /mol_type="genomic DNA"
 /sub_species="paratuberculosis"
 /db_xref="taxon:1770"
 /clone="mpcb61.15"
 /clone_1ib="Mycobacterium avium subsp. paratuberculosis
 DNA"

BASE COUNT 70 a 96 c 93 g 71 t
 ORIGIN

Query Match 5.0%; Score 63; DB 28; Length 330;
 Best Local Similarity 52.9%; Pred. No. 0.00016;
 Matches 135; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

OY 637 CTCGATGAGCGCGGAGAAACCCCTCGAATAAGCTGTGACGATGCTGTCAGCC 696
 Db 14 CTCGCGGCGCGGAGAAACCCCTCGAATAAGCTGTGACGATGCTGTCAGCC 73
 OY 697 GAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 756
 Db 74 GAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 133
 OY 757 GCTGCTGCGACCGATACCGATACCGATACCGATACCGATACCGATACCGATACCG 816
 Db 134 GATGCTGAGGATACCGATACCGATACCGATACCGATACCGATACCGATACCGATACCG 193
 OY 817 TCGCCGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGAG 876
 Db 194 CATCCCAAGAACTGAGCGGCTTAAAGCGAGACCTTCACTGCGCGGCGGATCGAA 253
 OY 877 GAGCGCTGCGGCTTC 891
 Db 254 GAAATGCTGCGCTAC 268

RESULT 4
 BZ548452 1127 bp DNA linear GSS 17-DEC-2002
 LOCUS BZ548452
 DEFINITION pac61-60_1124.s1 pac61-60 Pseudomonas aeruginosa genomic clone
 BZ548452
 accession pac61-60_1124, genomic survey sequence.
 VERSION BZ548452.1 GI:27152033
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 1 (bases 1 to 1127)
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
 Burns, J.L., Kaul, R. and Olsen, M.V.
 TITLE Whole-Genome-Sequence variation among multiple isolates of
 Pseudomonas aeruginosa library
 J. Bacteriol. (2002) In press
 COMMENT Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954

Fax: 2066857244
 Email: ckraymond@u.washington.edu
 Class: shotgun.
 Location/Qualifiers

FEATURES
 source 1..1127
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="1-60"
 /db_xref="taxon:287"
 /clone="pac61-60_1124"
 /clone_1ib="pac61-60"
 /note="Clinical isolate 1-60 whole genomic shotgun
 library."

BASE COUNT 197 a 379 c 340 g 210 t 1 others
 ORIGIN

Query Match 4.4%; Score 55.4; DB 29; Length 1127;
 Best Local Similarity 48.3%; Pred. No. 0.014;
 Matches 186; Conservative 0; Mismatches 196; Indels 3; Gaps 1;

OY 284 TGTTCGGGCTGCGCGGAGATCAGCGCTGCGGAGTCCGCAAGCTGCAACCGCTGTTA 343
 Db 252 TGTTCAGCTGACCGCGGAGCAATACCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 311
 OY 344 GGTCAAGCGGATGACCTGCTGCGCGGAGATACAGCGGATGACCGGATGACCGGATGAC 403
 Db 312 CCGCGCGGAGTGAAGCGCTGCAACCGGATATACAGGATCAGCGGATGATGCTG 371
 OY 404 ATGCTGCTCGGACAGAGAGTTCAGCTTGTGCGGATTTAGCGGAGGAAATCCGA 463
 Db 372 ACGCCATGCGCGCGCGGAGAGCGGATGAGCGGATGAGCGGATGAGCGGATGAGCGG 431
 OY 464 TGGCGGATGAGCGGCTGTTGAAGTTCGCGCGGAGTGAAGAGTTCGCTGCT 523
 Db 432 TCGCGGATTTTGAAGTTCGCTGCGGATTTCCGAGCGGAGCGGAGCGGAGCGGAGCGG 491
 OY 524 TCGGCTGCGGATGCGCGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 580
 Db 492 CCGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 551
 OY 581 AGACCAAGACCTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 640
 Db 552 AGGTGACTACTGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTG 611
 OY 641 ATGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665
 Db 612 ACAGCGGCTGTTAAGGCGGCGGA 636

RESULT 5
 CNS0091P 925 bp DNA linear GSS 03-JUN-1999
 LOCUS CNS0091P
 DEFINITION Drosophila melanogaster genome survey sequence T873 end of BAC #
 BACR19D16 of RPI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 accession AL053013
 version AL053013.1 GI:4934461
 keywords GSS.
 source Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 925)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information

COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu

FEATURES
source
1. 609
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3053"
/clone_lib="C. reinhardtii CC-1690, Stress II (normalized)
, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
XhoI; Stress condition II library, constructed by John
Davies and Jeffrey McDermott, combines cDNAs from CC-1690
cells grown to mid-log phase in TAP (NH4+ - containing)
and shifted to TAP - NO3- (24hrs); H2 production
conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
PolyA mRNA was purified from each sample, pooled and cDNA
synthesized. The cDNA was directionally cloned into lambda
Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
sites. pBluescript II SK- plasmids were excised from the
lambda Zap clones by superinfection with Exsist
(Stratagene) phage. The library was normalized using
method 4 described in Bonaldo et al., (1996) Genome
Research 6: 791-806."

BASE COUNT 119 a 200 c 213 g 77 t

ORIGIN

Query Match 4.1%; Score 52; DB 12; Length 609;
Best Local Similarity 48.9%; Pred. No. 0.073;
Matches 139; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

Qy 568 CAGCTCGATGAGAGACCAAGACCTGTCGCGTCCGTCACGAGGAGCGTCGCTGC 627
Db 32 CAGCTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 91
Qy 628 CATGACGTCCTGATGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 687
Db 92 CACGCTGCTGCTGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
Qy 688 CTTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 747
Db 152 CAGAACCGCGCTGGCCACTTGTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 211
Qy 748 GCGATTATCGCTGCTGCGACCGATCAAGATCTTATCGCGTGTGCTGCTAC 807
Db 212 GCGTTCCTCAACGCGGTACGAGAGACCATCTTCCGCGAGCGGCGGCGGCGGCGG 271
Qy 808 CTGCTCGGTCGCGCGGAGCGGCTGAGCTGTGTAAGCGGAGCC 851
Db 272 TTCCCGCGCGCTACGAGCGCTACATGACGAGGAGGAGGCGC 315

RESULT 8
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19D16 of RPCT-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013.1 GI:4934461
VERSION AL053013
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequences
BP 191 91006 EVRY cedex - FRANCE (E-mail: sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aron Mammeter in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCT-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. 925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACR19D16"
/clone_lib="RPCT-98"
/note="end : TET3"

BASE COUNT 120 a 61 c 61 g 172 t 511 others

ORIGIN

Query Match 4.1%; Score 51.8; DB 29; Length 925;
Best Local Similarity 14.5%; Pred. No. 0.092;
Matches 52; Conservative 159; Mismatches 148; Indels 0; Gaps 0;

Qy .530 CGGAGCTGCGCGCGCGCTGCGGCTGCTGTCGCCCGATGATGAGAGACCAAGA 589
Db 925 CSBSCSGSGSGBSS 866
Qy 590 CCTGTGTCGCTCCGTCAACGAGGCGCTGCTGCTCATGACTCTCATGAGCGGC 649
Db 865 SSGTSSACVKNMASSGCGCGGABCMCSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 806
Qy 650 GCAGAACCGCGTGAATAAGATGCTTGAAGAGCTGCTTCAAGCGGCGGCGGCA 709
Db 805 GGASASHSSSSACSS 746
Qy 710 GCAGCTGACGACGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 769
Db 745 ASASVSVAASSS 686
Qy 770 ATACGAGATCTTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 829
Db 685 TSMSCSSTSAASMASS 626
Qy 830 TCGAGCTGTGAGAGCGGCGGCGGCTCAAGAGAACGCGCTGATGAGAGTGTGCT 888
Db 625 SMSGSGSGSVSS 567

RESULT 9
BG488189 425 bp mRNA linear EST 27-MAR-2001
LOCUS BG488189
DEFINITION RH122_60_F03_b1_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA
Sequence.
ACCESSION BG488189
VERSION BG488189.1 GI:13469424
KEYWORDS EST.
SOURCE Sorghum propinquum
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

REFERENCE AUTHORS	TITLE JOURNAL COMMENT
Clade; Panicoidae; Andropogoneae; Sorghum. 1 (bases 1 to 425) Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt L. H.	An EST database from Sorghum: <i>Sorghum prostratum</i> rhizomes unpublished Contact: Cordonnier-Pratt MM cordonnier@uconnvm.uconn.edu

```

FEATURES
    source
        location/Qualifiers
            1..425
                /organism="Soygum propinquum"
                /mol_type="mRNA"
                /db_xref="taxon:132711"
                /clone_id="Rhizome2 (RAH22)"
                /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
zap II; Site 1: XhoI; Site 2: EcoRI. The library was made
from poly-A RNA in the cloning vector lambda zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT
    69 a      136 c      156 g      63 t      1 others
ORIGIN

```

RESULT 10	BZ569411/c	1170 bp	DNA	linear	GSS 17-DEC-2002
LOCUS					
DEFINITION					
ACCESSION	BZ569411				
VERSION	pac82-164_8374.x101				
KEYWORDS	pac82-164_8374, genomic survey sequence.				
SOURCE	BZ569411				
ORGANISM	BZ569411.1	GI:27203966			
	GSS.				
	<i>Pseudomonas aeruginosa</i>				
	<i>Pseudomonas aeruginosa</i>				
	Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
	<i>Pseudomonadaceae</i> ; <i>Pseudomonas</i> .				
	1 (bases 1 to 1170)				
REFERENCE	Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,				
AUTHORS	Burns, J.L., Kaul, R. and Olsen, M.V.				
TITLE	Whole-genome-Sequence variation among multiple isolates of				
JOURNAL	<i>Pseudomonas aeruginosa</i> library				
COMMENT	J. Bacteriol., (2002) In press				
	Contact: Chris K. Raymond				
	Genome Center				
	University of Washington				
	Box 352145, Seattle, WA 98105-2145, USA				
	Tel: 2062216954				

```

FEATURES
  source
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="2-164"
    /db_xref="taxon:287"
    /clone="pacs2-164_8374"
    /clone_1fb="pacs2-164"
    /note="clinical isolate 2-164 whole genomic shotgun
    library."
BASE COUNT
92 a      489 c      543 g      9 t      37 others
ORIGIN

```

[illegible]

RESULT 11
LOCUS CD045174 .
DEFINITION
ACCESSION CD045174
KEYWORD 512 bp mRNA linear EST 09-MAY-2001
SYNOPSIS psHB018xC17.6_183952 psHB: Infected hypocotyl soybean host: 48 hrs
post infection Phytophthora sojae cDNA clone sHB018C17 5, mRNA
sequence.
CD045174
CD045174.1 GI:30498767
EST.
Phytophthora sojae
Phytophthora sojae
Eukaryota: stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
1 (bases 1 to 512)
Tyler, B.M., Judelson, H.S., Gijzen, M., Dean, R.A. and Maugh, M.E.
USDA-IFRAs: Expression of Phytophthora sojae genes during infection
and propagation
unpublished
Contact: Tyler B
Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmt Tyler@vt.edu

PCR Primers
 FORWARD: BK reverse
 Plate: 018 row: C column: 17
 Seq primer: BK reverse
 High quality sequence stop: 512
 location/Qualifiers
 1
 512

BASE COUNT	105 a	. 159 c	182 g	66 t
ORIGIN				

Query Match	4.0%;	Score 50.2;	DB 14;	Length 512;
Best Local Similarity	45.5%;	Pred. No. 0.18;		
Matches 178; Conservative	0;	Mismatches 213;	Indels 0;	Gaps 0;

Qy 546 GCTCGGCGTGGGTTTGGTCCCAAGTGCATAGAGACCAAGACCCTGGTCGGTCCGT 605

Db 66 GCCCGTGGTGACTAGCCGCTCAAGATCCCGGTGAGGCCCAAGGCGGTGACCTCAAGAA 125

Qy 606 CACCGAGGGGCTCGCGCTCTGCATGACGTCTCGATGAGGCGCGCGAGAAACCCGCTCGA 665

Db 126 CGTCAAGATGAGCATGAACCCCTCTTCGAGATCCCGCTGGAGAGGCGCATCCGCTTCAA 185

Qy 666 AATGACGTTCTTGACGATCTGCTTCAAGCCGAGGCCCAACGGCAGACGCTGAGCAGAA 725

Db 186 GGGAGAAAGCTGGGCCCAACGAGGTGCTGGCCGCTGTCCATCGGCCCCCAAGAGAACCGGA 245

Qy 726 GGAGCTGGTCCGCGCTGTCGGTGGCATTAATGCTGTCGTCGACCCGATGCAAGATCTACT 785

Db 246 GACGCTGGCGACCGCGCTGGCCATAGGCGCGGACCCGCGCATTCACATCACACGAGCAT 305

Qy 786 TATCGCGTTCCGCTGTGCTCAACCTGTGCGGTCCGCCAGGCGCTCGAGCTGTGAAAGC 845

Db 306 GCGCACGAGACAGAGAGCTGCAGCGCGGTGGCCGATGAGTCAAGAGAGGTCCGTGGC 365

Qy 846 CGAGCCCCGGGCTCATAGAGAACGCGCTCGATGAGGTGCTCCGCTTTCGACAAATATCTCAAG 905

Db 366 CAAAGAGAGGCCGACAGCTGTCATCTGCGGCAAGAGCAGATCGACGCCGACCGCGCA 425

Qy 906 AATAGGAACTGTGCTTTTCGCCAGGCGAGAC 936

Db 426 GACGGGCCCCATGCTGGCGGCGCTGTGGAC 456

RESULT	12
LOCUS	CD045351
DEFINITION	CD045351 512 bp mRNA linear EST 09-MAY-2001 pBH8019xF20F.184338 pSBH: Infected hypocotyl soybean hose. 48 hrs post infection Phycophthora sojae cDNA clone SBH019F20 5', mRNA
ACCESSION	CD045351
VERSION	CD045351
KEYWORDS	Sequence.
SOURCE	EST.
ORGANISM	Phytophthora sojae
	Phytophthora sojae
	Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
	Phytophthora.
REFERENCE	1 (bases 1 to 512)
AUTHORS	Tyler B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
TITLE	USA-FRPS: Expression of Phytophthora sojae genes during infection and propagation
JOURNAL	unpublished

COMMENT Contact: Tyler B
m-1-2-1-4

1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtylex@vt.edu

```

FORWARD: BK reverse
Plate: 019 row: F column: 20
Seq primer: BK reverse
High quality sequence stop: 512

```

BASE COUNT	105 a	159 c	181 g	67 c
ORIGIN .				

```
Query Match      4.0%; Score 50.2; DB 14; Length 512;
Best Local Similarity 45.5%; Pred. No. 0.18;
Matches 178; Conservative 0; Mismatches 213; Indels 0; Gaps 0;
```

Qy	546	GCTCGGCGCTGGGTTTGGTGGCCCGACAGTGGATGATGAGAGACCAAGACCTTGGTGGCGTCCGT	605
Db	66	GCGCGTGGTGGACATACGCCGCTTCAGATTCGCGGTGAGGCCCAAGGGGCTTGACCTCAAGAA	125
Qy	606	CACCGAGGGGCTCCGCGCTGCTTCATGACGTCTCGATGAGCGGCGCAGGAATCCCGCTCGA	665
Db	126	CGTCAAGATGAGACATGAATCCCTTCTCGGAGATCCCGGTGAGGAGGCCATCCGCTTCAA	185
Qy	666	AAATGACCTCTTGAACGATGCTGCTTCAGGCGGAGGCCGACCGGACAGGCTAGACACGA	725
Db	186	GGAAGAGAGCTGGGCGACCGAGAGTGTGAGCGGTGTCATCGGCCCCCAAGCAAGACCGAA	245
Qy	726	GAGCTGGTTCGGGCTCGTGGGTGCATTATCGCTCTGTGCAACGGAATACCAACATCTACCT	785
Db	246	GACGCTGGCGCACGGGGCTGGCATGGGGCGGCAACCGCGGATCTCAATACCAACGGAACAT	305
Qy	786	TATCGCGTTGCTGTGCTCAACCTTGCTGCGGTGCGCCGAGGCGGCTCGAGCTGGTGAAGGC	845
Db	306	GCGCACGAGACAGGAGGCTGCAAGCCGCTGAGCGGTGACCAAGCTCTCAAGGAAGTCTGGGC	365
Qy	846	CGAGCGCCGGGCTCATGAGAGAAACGGGCTGATGAGGTGCTCCGCTTTCGAACAATATCTCAG	905
Db	366	CAAGGAGAGGCCGACGCTCGTATCTGGCGCAAGAGAGAGCATGACGCTCCGACGCCGCGCA	425
Qy	906	AATAGGAATGTCGCTTTCGCACGACGAGAC	936
Db	426	GACGGGCCCATGCTGGCCGGCTTGCTGGAC	456

RESULT 13	
LOCUS	CD260296
DEFINITION	CD260296 596 bp mRNA linear EST 23-MAY-2003
ACCESSION	CD260296
VERSION	CD260296.1
KEYWORDS	CD260296.1 GI:31048123
SOURCE	EST.
ORGANISM	Phytophthora sojae
	Phytophthora sojae
	BuNaryota; Stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.

REFERENCE 1 (bases 1 to 596)
AUTHORS Tyler, B.M., Judelson, H.S., Gijzen, M., Dean, R.A. and Maugh, M.E.
TITLE USDA-IPAFS: Expression of Phytophthora sojae genes during infection
JOURNAL and propagation
COMMENT Unpublished
Contact: Tyler B
Tyler Lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtylet@vt.edu
PCR Primers
FORWARD: BK reverse
Plate: 008 row: I column: 18
Seq primer: BK reverse
High quality sequence stop: 596.
Location/Qualifiers
1. 596
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sma008118"
/issue_type="mycelium"
/cell_line="P6497"
/dev_stage="mycelium"
/clone_1ib="sma: Phytophthora sojae grown in synthetic medium"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 119 a 190 c 212 g 75 t
ORIGIN

Query Match 4.0%; Score 50.2; DB 14; Length 596;
Best Local Similarity 45.5%; Pred. No. 0.19;
Matches 178; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 546 GCTCGGCTGGGTTGTGTCGCCAGGTGATGAGAGACCAAGACCTGTGCGCTCGT 605
DB 67 GCGCGTGTGACTAGCCGCTCAAGATCGCGTGAGAGCCCAAGGGGTGACTCAAGA 126
QY 606 CACCGAGGGGCTCGCGCTCTCCATGACCTCTCGATGAGCGCGCAGAACCCGCTCA 665
DB 127 CGTCAAGATGAGATGAACCCCTTCTGCGAGATCGCGTGAGAGAGCCATCGGCTCA 186
QY 666 AATGACGCTTGAAGATGCTGCTTCAAGCGGAGCGGAGCGGAGAGGCTGAGCAGAA 725
DB 187 GGAAGAAAGCTGGCCACGAGGTGCTGTCGATGAGCCCAAGAGAGCCAGGA 246
QY 726 GGAGCTGTGCGGCTCGTGGTGCGATTATCGCTGTGCGACCGATACAGATCTACCT 785
DB 247 GACGCTGGGCGACGGCGCTGCGCATGGGCGGAGCCGGGCGATCAGATCAGCAGCA 306
QY 786 TATCGGCTTGGCTGTGCTCAACTGCTGCGGTCGCCGAGGCGCTGAGCTGTGAAGC 845
DB 307 GCGCAGGAGCAGAGAGCTGACGCGCTGGCGCTGCGCAAGCTGCTCAAGAGAGTCTG 366
QY 846 CGAGCGCGGAGCTCATAGAGAGCGGCTCGATGAGGTCCTCGTTGACAATATCTCAG 905
DB 367 CAAGAGAGAGCGGAGCTGTATCTGCGGAGAGCAGAGATGAGCGGAGCGCGCGCA 426
QY 906 AATGAGAACTGTGCTTTCGCCAGGAGAGAC 936
DB 427 GACGGGCGCCATGCTGGCGGCGCTGTGAGAC 457

RESULT 14
CD043612 610 bp mRNA linear EST 09-MAY-2003
LOCUS CD043612
DEFINITION pSHB009XH01f_180418 pSHB: Infected hypocotyl soybean host. 48 hrs
post infection Phytophthora sojae cDNA clone sHB009XH01 5, mRNA
sequence.
ACCESSION CD043612
VERSION CD043612
KEYWORDS EST. GI:30497205

SOURCE
ORGANISM Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora

REFERENCE 1 (bases 1 to 610)
AUTHORS Tyler, B.M., Judelson, H.S., Gijzen, M., Dean, R.A. and Maugh, M.E.
TITLE USDA-IPAFS: Expression of Phytophthora sojae genes during infection
JOURNAL and propagation
COMMENT Unpublished
Contact: Tyler B
Tyler Lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtylet@vt.edu
PCR Primers
FORWARD: BK reverse
Plate: 009 row: H column: 01
Seq primer: BK reverse
High quality sequence stop: 610.
Location/Qualifiers
1. 610
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sHB009H01"
/issue_type="infected host tissue"
/cell_line="P6497"
/dev_stage="48 hour post infection"
/clone_1ib="pSHB: Infected hypocotyl soybean host. 48 hrs post infection"
/note="vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; USDA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation."

BASE COUNT 121 a 199 c 215 g 75 t
ORIGIN

Query Match 4.0%; Score 50.2; DB 14; Length 610;
Best Local Similarity 45.5%; Pred. No. 0.19;
Matches 178; Conservative 0; Mismatches 213; Indels 0; Gaps 0;

QY 546 GCTCGGCTGGGTTGTGTCGCCAGGTGATGAGAGACCAAGACCTGTGCGCTCGT 605
DB 69 GCGCGTGTGACTAGCCGCTCAAGATCGCGTGAGAGCCCAAGGGGTGACTCAAGA 128
QY 606 CACCGAGGGGCTCGCGCTCTCCATGACCTCTCGATGAGCGGCGCAGAAACCGCTCA 665
DB 129 CGTCAAGATGAGATGAACCCCTTCTGCGAGATCGCGTGAGAGAGGCGATCGGCTCA 188
QY 666 AATGACGCTTGAAGATGCTGCTTCAAGCGGAGCGGAGCGGAGAGGCTGAGCAGAA 725
DB 189 GGAAGAAAGCTGGCCACGAGGTGCTGTCGATGAGCCCAAGAGAGCCAGGA 248
QY 726 GGAGCTGTGCGGCTCGTGGTGCGATTATGCTGTGCGACCGATACAGATCTACCT 785
DB 249 GACGCTGGGCGACGGGCTGCGCATGGGCGGAGCGGCGATTCAGATCAGCAGCA 308
QY 786 TATCGGCTTGGCTGTGCTCAACTGCTGCGGTCGCCGAGGCGCTGAGCTGTGAAGC 845
DB 309 GCGCAGGAGCAGAGAGCTGACGCGCTGGCGCTGCGCAAGCTGCTCAAGAGAGTGTG 368
QY 846 CGAGCGCGGAGCTCATAGAGAGCGGCTCGATGAGGTCCTCGTTGACAATATCTCAG 905
DB 369 CAAGAGAGAGCGGAGCTGTATCTGCGGAGAGCAGAGATGAGCGGAGCGCGCGCA 428
QY 906 AATGAGAACTGTGCTTTCGCCAGGAGAGAC 936
DB 429 GACGGGCGCCATGCTGGCGGCGCTGTGAGAC 459

RESULT 15
CD262676 623 bp mRNA linear EST 23-MAY-2003
LOCUS CD262676

Search completed: October 3, 2003, 17:48:11
Job time : 2496.11 secs

DEFINITION psmA018K21f.191994 sMA: Phytophthora sojae grown in synthetic
medium Phytophthora sojae cDNA clone sMA018K21 5, mRNA sequence.
ACCESSION CD262676
VERSION CD262676.1 GI:31050503
KEYWORDS EST.
SOURCE Phytophthora sojae
ORGANISM Phytophthora sojae
Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
Phytophthora.
REFERENCE 1 (bases 1 to 623)
AUTHORS Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
TITLE USDA-IRAFs: Expression of Phytophthora sojae genes during infection
and propagation
JOURNAL Unpublished
COMMENT Contact: Tyler B
Tyler lab
VBI
1880 Pratt Dr., Blacksburg, VA 24061, USA
Tel: 540-231-7318
Email: bmtyle@vt.edu
PCR PRIMERs
FORWARD: BK reverse
Place: 018 row: K column: 21
Seq primer: BK reverse
High quality sequence stop: 623.
Location/Qualifiers
1..623
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="sMA018K21"
/isue_type="mycelium"
/cell_line="P6497"
/dev_stage="mycelium"
/clone_1fb="sMA: Phytophthora sojae grown in synthetic
medium"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 123 a 206 c 217 g 77 t
ORIGIN
Query Match 4.0%; Score 50.2; DB 14; Length 623;
Best Local Similarity 45.5%; Pred. No. 0.19;
Matches 178; Conservative 0; Mismatches 213; Indels 0; Gaps 0;
QY 546 GCTGGCGGTGGTGGTGGCCAGTGGATAGAGACCAAGACCTGGTGGTGGT 605
DB 59 GCGCTGTGTGACTACGCCGCTCAAGATCCGCTGAGGCCAAGGCGTGGACTCAAGAA 118
QY 606 CACCGAGGGGCTCGGCGTGTCTCATGACGTCCTCGATGAGCGGCGCAGGAACTCCGTCGA 665
DB 119 CGTCAAGATGACATGAACCTCTTCTGGAGATCCCGCTGGAGAGGCCATCCGCTCA 178
QY 666 AAATGACGCTTTGACGATGCTGCTTCAAGGCCGAGGCCGACGCGCAGAGCTTGACAGAA 725
DB 179 GGAGAGAAAGCTGGCCACCGAGGTGCTGGCCGCTGTCATCGGCCCCCAAGCAAGCAGGA 238
QY 726 GGAGCTGTGGGCTCGGCGTGTGCGATATATGCTGCTGGCACCGATACGATCTAAGT 785
DB 239 GACGCTGTGCGACGGCGCTGGCCATGAGCGCGGACCGGATTCATCAACCAAGACAT 298
QY 786 TATCCGCTGTGCTGTCAACCTGCTGGGATCGCCCGAGGCGCTCGAGCTGTGAAGGC 845
DB 299 GCGCACGAGACGAGAGCTGCAAGCCGCTGGCCGCTGCAAGCTGTCAAGAGAGTGTGGC 358
QY 846 CGAGCCCGGCTCATGAGGAACGCGCTCGATGAGTGTCTCCGCTTGACATATATCTCAG 905
DB 359 CAAGAGAGGCGCGAGCTGTGATCTGCGGCAAGCAGAGCATCGACCGCAGCGCGCA 418
QY 906 AATGAGACTGTGCGTTTGGCCAGGAGAC 936
DB 419 GACGGCCCATGTGGCCGCTGCTGAGC 449

THIS PAGE BLANK (USPTO)

Db 507 -----GGAGCGCCAGCGCGGAATCTCTCGCCGAGAGAGCCCAAGCCCTTGCCGATG 559
 Qy 614 GGCTGCGCTGCTGCATGACGTCCTCGATGACGGGCGAGAACCCGCTCGAAAAATGAG 673
 Db 560 CCGAGGTGACATACCTGCGCGCTGCTCGAGGCCAAGGCGCGGCGAGCGGCGAGACG 619
 Qy 674 TCTTGACGATGCTGCTTACGCGCGGAGCGGCGAGCGGCGAGCGGCGAGCGGCGAG 733
 Db 620 TCTTACGCGCGGCTGCGAGCGCGCGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGAG 679
 Qy 734 TCGCGCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 793
 Db 680 TCTTACGCGCGGCGAGCGGCGCGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGAGCGG 739
 Qy 794 TCGCTGCTGCTCAACCTGCTGCGCGCGCGCGAGCGGCGAGCGGCGAGCGGCGAGCGG 853
 Db 740 AGCGCGCTGCTCAACCTGCTGCGCGCGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGG 799
 Qy 854 GGCTCATGAGAACGGGCTCGATGAGTGTCTCGCTTGCACAAATATCTTCAAAATGAGA 913
 Db 800 AACTCTGCGCCAAACGCGATGAGAACTGCTGCGCGAGCGGCGAGCGGCGAGCGGCGG 859
 Qy 914 CTGTGCGCTTGCAGAGCGAGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGG 973
 Db 860 TGTGCGCTTGCAGAGCGAGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGG 919
 Qy 974 TGTGCTTCTCTGATCCGAGCGGCGGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGG 1033
 Db 920 AACTCTGCTGCTCAACCTGCTGCGCGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGG 979
 Qy 1034 TGTGCTGATGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGAGCGGCGG 1093
 Db 980 GCGTCACTCAACCGCGCAACCGATGCGCTGCTGCGCGAGCGGCGAGCGGCGAGCGG 1039
 Qy 1094 GCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1153
 Db 1040 GCGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1099
 Qy 1154 GGTTCGCGAGATGAAGCTGAAAGAACTCCGCTG 1188
 Db 1100 GCTTCCCGACCTCCAGTGGCGGCTGCCACGCG 1134
 RESULT 2
 PCT-US03-18787-2/c
 ; Sequence 2, Application PC/TUS0318787
 ; GENERAL INFORMATION:
 ; APPLICANT: KOSAN BIOSCIENCES
 ; APPLICANT: Hutchinson, Richard C.
 ; APPLICANT: Reid, Ralph C.
 ; APPLICANT: Hu, Zhihao
 ; APPLICANT: Rascher, Andreas
 ; APPLICANT: Schirmer, Andreas
 ; APPLICANT: McDaniel, Robert
 ; TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
 ; TITLE OF INVENTION: PRO-GEGLANAMICIN PRODUCING POLYMERASES AND
 ; FILE REFERENCE: 30622009740
 ; CURRENT APPLICATION NUMBER: PCT/US03/18787
 ; CURRENT FILING DATE: 2003-06-13
 ; PRIOR APPLICATION NUMBER: US 60/389,255
 ; PRIOR FILING DATE: 2002-06-14
 ; PRIOR APPLICATION NUMBER: US 60/393,929
 ; PRIOR FILING DATE: 2002-07-03
 ; PRIOR APPLICATION NUMBER: US 60/395,275
 ; PRIOR FILING DATE: 2002-07-12
 ; PRIOR APPLICATION NUMBER: US 10/212,962
 ; PRIOR FILING DATE: 2002-08-05
 ; PRIOR APPLICATION NUMBER: US 60/415,326
 ; PRIOR FILING DATE: 2002-09-30
 ; PRIOR APPLICATION NUMBER: US 60/420,820
 ; PRIOR FILING DATE: 2002-10-24
 ; PRIOR APPLICATION NUMBER: US 60/433,130

; PRIOR FILING DATE: 2002-12-13
 ; NUMBER OF SEQ ID NOS: 153
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2
 ; LENGTH: 86941
 ; TYPE: DNA
 ; ORGANISM: Streptomyces hygroscopicus
 PCT-US03-18787-2
 Query Match 10.3%; Score 130.2; DB 1; Length 86941;
 Best Local Similarity 50.4%; Pred. No. 1.1e-24;
 Matches 448; Conservative 0; Mismatches 408; Indels 33; Gaps 4;
 Qy 303 GGATACGCTGGGGTCCGCAAGCTGTACACCCGCTTTACGTACGCGCGCATGACCT 362
 Db 82492 GGAGCAGCCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 82433
 Qy 363 GCTGCGCGCGCAATATACAGCGACCGTGCAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 422
 Db 82432 GATGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 82373
 Qy 423 ---GGAGTTGACGCTTGTGCGGATTTACGCGAGGAAATCCGATGCGCGCATGAGCGC 479
 Db 82372 CCGCGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 82313
 Qy 480 TCTGTTGAAGTTCGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
 Db 82312 GATGCTGCGGCGCTTGTGAGCGCGGATGATGATGATGATGATGATGATGATGATGATGAT 82268
 Qy 540 GCGCGCGCTGCGCGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599
 Db 82267 GTCGAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 82208
 Qy 600 GTCGCTCAACGAGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
 Db 82207 CATGCTGAGCTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 82148
 Qy 660 GCTGCAAAATGACGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 719
 Db 82147 GCTCA-----GCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 82103
 Qy 720 CACGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 779
 Db 82102 CAGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 82043
 Qy 780 CTACCTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 839
 Db 82042 CAACATGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 81983
 Qy 840 GAAGCGCGACCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 899
 Db 81982 GCGGCGCGATGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 81923
 Qy 900 CCTCAGATGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 959
 Db 81922 CCGTCAAGAGAGCAACGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81863
 Qy 960 CAAGAGGCGAGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019
 Db 81862 CCGGCGCGGAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 81803
 Qy 1020 CTCAGAGCGCAAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1079
 Db 81802 GAGCGCGCAAGCTGAGCTGAGCGCGCGCAACCGCGCGAGATGATGATGATGATGATGATGAT 81743
 Qy 1080 AGGCGCGCATGCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1139
 Db 81742 CCGGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 81683
 Qy 1140 CACCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1188
 Db 81682 GACCTGCTGCGCGCGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81634

RESULT 3
PCT-US03-19069-2/c
Sequence 2, Application PC/TUS0319069
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES
APPLICANT: Hutchinson, Richard C.
APPLICANT: Reid, Ralph C.
APPLICANT: Hu, Zhihao
APPLICANT: Raescher, Andreas
APPLICANT: Schlimer, Andreas
APPLICANT: McDaniel, Robert
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
TITLE OF INVENTION: PRO-GEIDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
FILE REFERENCE: 300622009740
CURRENT APPLICATION NUMBER: PCT/US03/19069
CURRENT FILING DATE: 2003-06-16
PRIOR APPLICATION NUMBER: US 60/389,255
PRIOR FILING DATE: 2002-06-14
PRIOR APPLICATION NUMBER: US 60/393,929
PRIOR FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/395,275
PRIOR FILING DATE: 2002-07-12
PRIOR APPLICATION NUMBER: US 10/212,962
PRIOR FILING DATE: 2002-08-05
PRIOR APPLICATION NUMBER: US 60/415,326
PRIOR FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: US 60/420,820
PRIOR FILING DATE: 2002-10-24
PRIOR APPLICATION NUMBER: US 60/433,130
PRIOR FILING DATE: 2002-12-13
NUMBER OF SEQ ID NOS: 153
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 86941
TYPE: DNA
ORGANISM: Streptomyces hygroscopicus
PCT-US03-19069-2

Query Match 10.3%; Score 130.2; DB 1; Length 86941;
Best Local Similarity 50.4%; Pred. No. 1,1e-24;
Matches 448; Conservative 0; Mismatches 408; Indels 33; Gaps 4;
QY 303 GGATCAGCCTGCGGCGCCCAAGCTGCTCAACCCGCTGTTAGTCAGCGCCGATGACT 362
DB 82432 GGAGACGCGCGGATCCCGCGATGCTCAACCGCACTTCACTCCGCGCAGGCG 82433
QY 363 GCTGCGCCCGAATATACAGCGCAGCTGCTGATGCTGCTGCTCGACCAAGA 422
DB 82432 GATGCGCGCGATGATCCAGGGGATGCTGACGCGCTCTCGGACCGGCTGATGCGCCAGG 82373
QY 423 ---GAGATTCAAGCTTGTGCGGGAATTACGCGGAGGAATCCGATGCCGCGATAGCGC 479
DB 82372 CCGCGCGCGCAGCTGATGCGCGGCTTCCCGCTGCGCTCCAGGTGATGCGCGT 82313
QY 480 TCTGTTGAAGGTTCCGCGCGAGTGAAGAGAAAGTTCCGCTGCTTCCGCTGCGGCACTGC 539
DB 82312 GATGCTGCGGCTTCTGACCGCGGCTTCCGCGAGTTCCAGCG---GCG 82268
QY 540 GCGCGCGCTGCGCGGCTGCTGCTGCGCGGCTGCTGATGAGAGACCAAGACCGCTGCTGC 599
DB 82267 GTCGCGGCGCTGATGAGCTTCAACGCGCTGCGCGGAGAGAGAGGCGCGCTGCGGCT 82208
QY 600 GTCCGCTACCGAGGCGCTGCGCTGCTGCTGATGAGCTTCTGATGAGCGCGCGAGAACCC 659
DB 82207 CATGCTGAGCTACGAGCCCGATGTCGCGCGCAGGCGCGCGCGAGAGATCT 82148
QY 660 GCTCGAATATGACGCTTGAAGATGCTGCTTCAAGCCGAGCGCGAGAGAGGCTGAG 719
DB 82147 GCTCA-----GCGGCTCATGCTGATGAGACGAGCGCGG--GCGGCTCATC 82103
QY 720 CACGAAGAGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779

DB 82102 CCAGACGAGCTGATGCGCACCGCCCTGCTGCTGCTGCGCGCGCACGAGCACCGCC 82043
QY 780 CTACCTTATGCGGTTGCTGCTGCTCAACCTGCTGCGCGCGCGCGCTGAGCTGCT 839
DB 82042 CAACATGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81983
QY 840 GAAGCGCGAGCGCGGCTGATGAGGAAGCGCGCTGATGAGGCTGCTGCTGCTGCTGCT 899
DB 81982 GCGGCGGATGCGCGGCTGATGAGGCAAGCGGCTGAGCAACTGCTGCTGCTGCTGCTGCT 81923
QY 900 CTTGAAATGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
DB 81922 GCTGAGGAAGGCAACCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81863
QY 960 CAAGAAAGGAGATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019
DB 81862 CCGCGCGGTAAGGAGGATGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81803
QY 1020 CTGAGCGCAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079
DB 81802 GAGCGCCACGAGCTGAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCT 81743
QY 1080 AGGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139
DB 81742 CGGCGTGAACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 81683
QY 1140 CACCATCTTCCGTTAGTTCCTGCGAGATGAGCTGAGAACTCCCGG 1188
DB 81682 GACCTGCTGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGAG 81634

RESULT 4
PCT-US03-18787-1/c
Sequence 1, Application PC/TUS0318787
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES
APPLICANT: Hutchinson, Richard C.
APPLICANT: Reid, Ralph C.
APPLICANT: Hu, Zhihao
APPLICANT: Raescher, Andreas
APPLICANT: Schlimer, Andreas
APPLICANT: McDaniel, Robert
TITLE OF INVENTION: RECOMBINANT POLYNUCLEOTIDES ENCODING
TITLE OF INVENTION: PRO-GEIDANAMYCIN PRODUCING POLYKETIDE SYNTHASES AND
FILE REFERENCE: 300622009740
CURRENT APPLICATION NUMBER: PCT/US03/18787
CURRENT FILING DATE: 2003-06-13
PRIOR APPLICATION NUMBER: US 60/389,255
PRIOR FILING DATE: 2002-06-14
PRIOR APPLICATION NUMBER: US 60/393,929
PRIOR FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/395,275
PRIOR FILING DATE: 2002-07-12
PRIOR APPLICATION NUMBER: US 10/212,962
PRIOR FILING DATE: 2002-08-05
PRIOR APPLICATION NUMBER: US 60/415,326
PRIOR FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: US 60/420,820
PRIOR FILING DATE: 2002-10-24
PRIOR APPLICATION NUMBER: US 60/433,130
PRIOR FILING DATE: 2002-12-13
NUMBER OF SEQ ID NOS: 153
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1
LENGTH: 85692
TYPE: DNA
ORGANISM: Streptomyces hygroscopicus
PCT-US03-18787-1
Query Match 9.9%; Score 124.6; DB 1; Length 85692;
Best Local Similarity 48.0%; Pred. No. 3.5e-23;


```
Qy 553 GTGGTTGTTGTCCTCCAGCTGATGAGAGAACCAACCTGTGCTGCTCCGTACCCGAG 612
Db 3746 CCG-----ACCGACCCCGGACCGCGCTGTATAGAGGCCCGCTTGAC 3705
Qy 613 GGGCTCGCGCTGCTCATGAGAGTCTCTGATGAGCGCGCCAGGAACCCGCTGAAATATAC 672
Db 3704 GGGCTCCAGACTATATAGCCGAGGTCTG---CCCATTAAGCGGAACAGGGAGGAC 3648
Qy 673 GTCTTGAGATGCTGCTTCAAGCCGAGCGACGAGCAGCAGCTGAGCAGCAAGAGCTG 732
Db 3647 GACCTGCTCAGCCGCTGTGTGTGCGCCGACCGCAGCACTCTCCAGAGAGAGCTG 3588
Qy 733 GTCCGCTGTGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 792
Db 3587 GACTGTGATGATCTTCCAGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3528
Qy 793 TTGCTGTGCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 852
Db 3527 ACCGCTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3468
Qy 853 GAGCTCATGAGGAAACGCTGCTGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 912
Db 3467 GAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3408
Qy 913 ACTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 972
Db 3407 ACTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3348
Qy 973 ATGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1032
Db 3347 TCGGTGATCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3288
Qy 1033 GTGTTGATGTCGACGAGGACACGCGGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1092
Db 3287 ACCCTGACCTGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3228
Qy 1093 TGCCCCGGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1152
Db 3227 TGCCCCGGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3168
Qy 1153 AGCTTCCCCGAGATGACTG 1173
Db 3167 CGCTCTCCCGAGCGCACCCTG 3147

RESULT 6
US-10-297-465B-1/c
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Secubal, Joao
; APPLICANT: Mediana, Joao
; TITLE OF INVENTION: Isolated Genome of Xylella fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; PRIOR FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

Query Match 5.4%; Score 68.6; DB 6; Length 2731748;
Best Local Similarity 46.2%; Pred. No. 1.4e-07;
Matches 264; Conservative 0; Mismatches 304; Indels 3; Gaps 1;
```

```
Qy 602 CCGTACCGAGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 661
Db 38855 CTGCTACGAAAGACTGCGCAATATTTAGAGTGTATGACTACGCGCACCATC 388496
Qy 662 TCGAAATAGAGCTTTAGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721
Db 388495 CTGAACTGACTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 388436
Qy 722 CGAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781
Db 388435 ATGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 388376
Qy 782 ACTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 841
Db 388375 ATATGATGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 388316
Qy 842 AGCGCGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 901
Db 388315 AGAGTGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 388256
Qy 902 TCAGATAGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 961
Db 388255 TACAGTTT---ACGATGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 388199
Qy 962 AGAAGGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1021
Db 388198 CTGCTGAGACAGTATGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 388139
Qy 1022 CCAGCCAGAGCTTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1081
Db 388138 CTATCTCCGACACAGCTGATATTTATTTATTTATTTATTTATTTATTTATTTATTT 388079
Qy 1082 GCGCCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1141
Db 38078 GCATTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 388019
Qy 1142 CCATCTTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1172
Db 388018 CGCTTTGAACGCTACCGCATTTAAGCT 387988

RESULT 7
US-10-297-465B-1
; Sequence 1, Application US/10297465B
; GENERAL INFORMATION:
; APPLICANT: Simpson, Andrew
; APPLICANT: Reinach, Fernando
; APPLICANT: Secubal, Joao
; APPLICANT: Mediana, Joao
; TITLE OF INVENTION: Isolated Genome of Xylella fastidiosa and Uses Thereof
; FILE REFERENCE: FAPESP 202 US (10213376)
; CURRENT APPLICATION NUMBER: US/10/297,465B
; PRIOR FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: PCT/IB01/01618
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/209,906
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2731748
; TYPE: DNA
; ORGANISM: Xylella fastidiosa
US-10-297-465B-1

Query Match 4.4%; Score 55.8; DB 6; Length 2731748;
Best Local Similarity 45.4%; Pred. No. 0.0004;
Matches 241; Conservative 0; Mismatches 287; Indels 3; Gaps 1;
```



```
Db 18269 GGATGAGCTGCTCTGAGCTTCGCCAAGAGCTTCGACCGCTGCGCGCTGC 18328
QY 698 AGCCGACGCGAGCAGCTGAGCAGAAAGAGCTGTCGCGCTCGTGAGTGCGATTATCG 757
Db 18329 TGGGGGAGGGGGCGCGGTTTCGAGAGATGGGCAACGACATCCGTGACCGGACGAGG 18388
QY 758 CTGCTGGACCGATACACGATCTAATTATGCTGCTGCTGCTCAACCTGTCGCGT 817
Db 18389 TCGCGCGCGCGATCCCGCGCTCACCTACCGCGCTTGACCTGCTCGACCGCGAGCGG 18448
QY 818 CG 819
Db 18449 CG 18450

RESULT 10
US-10-084-846A-77
; Sequence 77, Application US/10084846A
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TRETZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO: 77
; LENGTH: 972
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avilABCI dna: partial sequence of coding strand 1; nucleotide 1
; US-10-084-846A-77

Query Match 3.9%; Score 49.4; DB 6; Length 972;
Best Local Similarity 46.5%; Pred. No. 0.0017;
Matches 195; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 288 CGGGCTGCCCGCGGAGATCACGCTCGGAGTCCGAACTCGTCAACCGTGTTTACGTC 347
Db 504 CGGCTCTGACCCGCGCAACCGCAAGAGGTCTGAAAGTGTCCGTGATGATGACGA 563
QY 348 AGCGCCATTCAGCTGCTGCGCGCGGAATACAGGCGACCGTGCACAGCTGCTGATGC 407
Db 564 GGGGACACCGGCTGCTCAACGACAGTACCTGAGAGAGGCGACCGGCTGCGGACCA 623
QY 408 TCGCTCCGAGCAAGAGATTGAGCTTGTGCGGAGTAAAGCGAGAGGATCCGATGCG 467
Db 624 GATCTCGGTGATGACGCGGCGCGGCTGCTGAGGAGCAACCCGAGACCTTGAATGC 683
QY 468 CGCGATACAGCGCTGTTGAAGTTCCGCGCGAGTGTGACGAAATTCCTGCGCTTGG 527
Db 684 CTTGCTGGGCGCGACCGGATGTTGATGCTCTGCGCGGACCGAGAGAC---CTGTGCGC 740
QY 528 CTCGCGGAGTGGCGCGCGCTGCGGCTGTTGTTGCCCCGAGTTCGATGAGAGACCA 587
Db 741 CGCGCGGCTAGGTGCGCGCTTCAACGAGGACCGAGCGGAGGAGCGCGGAGCGCG 800
QY 588 GACCTGATGCGCTGCTGACCGAGGAGGCTCGCGCTGCTCAATGACGCTCTGATGAGCG 647
Db 801 CCGTATGACGCGCGGATGACGAGACCGGCTGCGGACCTCGTGAATGCTGCACTGCT 860
QY 648 GCGCAGAAACCGCTGCAAAATGACGCTTGAACGATGCTTCAAGCGCGAGCGGACG 706
Db 861 GCGCGGAGAGGCGACCGGCTGCTGATGTGTGCTGCGCGCGCGCGGACGCTGATGAGG 919
```

```
RESULT 11
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TRETZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO: 1
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; US-10-084-846A-1

Query Match 3.9%; Score 49.4; DB 6; Length 59816;
Best Local Similarity 46.5%; Pred. No. 0.0063;
Matches 195; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 288 CGGGCTGCCCGCGGAGATCACGCTCGGAGTCCGAACTCGTCAACCGTGTTTACGTC 347
Db 34691 CGGCTGACCCGCGGCAACCGCAAGAGTCTGGAAGATGTCGTTGATGATGACGA 34750
QY 348 AGCGCCATTCAGCTGCTGCGCGCGGAATACAGGCGACCGTTCAGCACTGCTGATGC 407
Db 34751 GGGGACACCGCTGCTGCTCACACGACGATACCTGAGAGGCGCGACCGCTGCGGACCA 34810
QY 408 TCGCTCCGAGCAAGAGAGTTCGACGTTGTGCGGAGTAAAGCGGAGGAAATCCGATGCG 467
Db 34811 GATCTCGGTGATGACGCGGCGCGGCGGCTGTCGAGGCGACCCCGAGACCTGAAATC 34870
QY 468 CGCGATACAGCGCTGTTGAAGTTCGCGCGGAGTGTGACGAAATTCCTGCTGCTGCG 527
Db 34871 CTTGCTGGGCGGCGACCGGATGTTGATGACCTGCGGCGCACCGAGAC---CTGTGCGC 34927
QY 528 CTCGCGGATGCGCGCGCTGCGGCTGCGGCTGTTGATGCTCCAGAGTTCGATGAGAGACCA 587
Db 34928 CGCGCGGCTGAGGTGCGCGCTTACCGGAGACCGAGCTGAGAGCGCGGAGCGCG 34987
QY 588 GACCTGATGCGCTGCTGACCGAGGAGGCTGCGGCTGCTCATGACGCTCTGATGAGCG 647
Db 34988 CCGTATGACGCGCGGATGACCGGAGCGGCTGCGGAGCTCTGAGATGCTGCACTGCT 35047
QY 648 GCGCAGAAACCGCTGCAAAATGACGCTTGAACGATGCTTGAAGCGGAGCGGACG 706
Db 35048 GCGCGGAGAGGCGACCGGCTGCTGATGTGTGCTGCGCGCGCGCGGACGCTGATGAGG 35106

RESULT 12
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TRETZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
```

PRIOR APPLICATION NUMBER: DE 101 09 166.4
PRIOR FILING DATE: 2001-02-25
NUMBER OF SEQ ID NOS: 120
SOFTWARE: Patent Ver. 3.2
SEQ ID NO 2
LENGTH: 59816
TYPE: DNA
ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Query Match 3.9%; Score 49.4; DB 6; Length 59816;
Best Local Similarity 46.5%; Pred. No. 0.0063;
Matches 195; Conservative 0; Mismatches 221; Indels 3; Gaps 1;

QY 288 CGGCTGCGCCGAGAGATCAAGCTGGGCGCAAGCTGTCACCCGCTTTAAGTC 347
DB 25126 CGGCTGCGCCGAGAGATCAAGCTGGGCGCAAGCTGTCACCCGCTTTAAGTC 25067
QY 348 AGCGCGCATGACCTGCGCGCGCGAATAACAGCCACCGTCGACCTGCTGATGC 407
DB 25066 GGGCACCACTGCTGCTGCTCAACGCACTGCGAGAGCCGACCGCTGCGCAACA 25007
QY 408 TCGCTCCGACAAAGAGATTCAGCTTGTGCGGATTAACGCGAGGAATCCGATGCG 467
DB 25006 GATCTGATGATGACGCGGAGCGCGGAGTGTGAGGGAACCCCGAGACCTGAAGTC 24947
QY 468 CGGATCAAGCTCTGTTGAAGTTCCGCGCGAGTGTGACGAGAATTCGCTGCGG 527
DB 24946 CCGTGGGCGCGGAGACCGGATGCTGTGACCTGCGCGCGCACGAGAC---CTGTGCGC 24890
QY 528 CTGCGGACATGCGCGCGCGCTGCGGCTTGTGAGCCCGAGGATGATGAGACGAA 587
DB 24889 CGCGCGCTCAAGTGTGCTGCTTCAAGGAGACCGAGCGAGTGAAGCGCGAGCCCG 24830
QY 588 GACCTGTGCGCTGCTGCTGACGAGGCGCTGCGCTGCTCAATGACGCTCTGATGAGCG 647
DB 24829 CGTATCAAGCGCGCGTACGAGCGAGGCGCGGCGGCGGCTGTGAGATGTGACGCTGCT 24770
QY 648 GCGGAGAAACCGCTGAAATGACGCTTGAAGATGCTGCTTCAAGCCGAGGCGGAGC 706
DB 24769 GCGCGGAGCGGCGCAAGCGGTGTGATGTGCTGCGCGCGCGCGGCGGCGGCTGATGACG 24711

RESULT 13
PCT-US03-11231-190
Sequence 190, Application PC/TUS0311231
GENERAL INFORMATION:
APPLICANT: Corixa Corporation
APPLICANT: Day, Craig H.
APPLICANT: Hosken, Nancy A.
APPLICANT: Parsons, Joseph M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
FILE REFERENCE: 210121.53801PC
CURRENT APPLICATION NUMBER: PCT/US03/11231
CURRENT FILING DATE: 2003-04-09
NUMBER OF SEQ ID NOS: 267
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 190
LENGTH: 9369
TYPE: DNA
ORGANISM: HSV2
PCT-US03-11231-190

Query Match 3.9%; Score 48.8; DB 1; Length 9369;
Best Local Similarity 46.1%; Pred. No. 0.005;
Matches 201; Conservative 0; Mismatches 232; Indels 3; Gaps 1;

QY 313 CGGCTCGCAAGCTGCTCAACCGCTGTTAGTCAAGCGGCAATGACCTGCTGCGCGC 372
DB 4489 CGCGATGCGCTGACCGCATCCCGGCGGAGTGCACCGGCGCGCTCGCGCGCGCTGCG 4548
QY 373 GAAATACAGCGACCGTCAACGCTGTGATGCTGTGCTCGGAGCAAGAGAGTTGAC 432

DB 4549 GACCTGGCGCGCGCGCGGAGATATCCGCGCGCGGAGATGACCCGAGGCCACCGCG 4608
QY 433 GTTGTGCGGAGATTACCGGAGGAAATCCGATGCGCGGATGACGCTTGTGAAGTT 492
DB 4609 GAGCTGTGTCGAGACCGGAGGATCACTGCTGTGCTCCAGCAGAGAGCGCTGTGCGTGC 4668
QY 433 CGGCGGATGATGAGAGAAATTCGCTGCTTCCGCTGCGGAGATGACCGCGCGCTGCGG 552
DB 4669 GCGGCGGCTTGAAGACCGGATGACGAGCCCGGAGAAACAGCGCTGACGAGATAT 4728
QY 553 GTGCTTTGTCGCGC---CAGTTCATGAGAGAGACCAAGACCTGCTGCTGCTGAC 609
DB 4729 GTGCTTTCATGACCGCGGAGAGACAGCGCCGAGATCAAGAGAGAGAGTGTGCGCGCAA 4788
QY 610 GAGGAGCTGCGCTGCTCAATGACCTCTGATGAGCGCGGAGAGACCGCTGCAAAAT 669
DB 4789 CAGCAGCGCGCGGAGAGACGAGCGGATCATGCGGCGCTGCGGAGAGCGCTGCGCGC 4848
QY 670 GACGCTTGAAGATGCTGCTTCAAGCGGAGCGGCGGAGCGGAGAGAGAGAG 729
DB 4849 CGCGAGCGCGCGCGCGGAGATGAGCGGAGCGGAGCGGAGCGGAGAGAGAGAG 4908
QY 730 CTGCTGCGCTGCTGCTG 745
DB 4909 GTGCTGCGCTGCGCG 4924

RESULT 14
PCT-US03-11231-247
Sequence 247, Application PC/TUS0311231
GENERAL INFORMATION:
APPLICANT: Corixa Corporation
APPLICANT: Day, Craig H.
APPLICANT: Hosken, Nancy A.
APPLICANT: Parsons, Joseph M.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
FILE REFERENCE: 210121.53801PC
CURRENT APPLICATION NUMBER: PCT/US03/11231
CURRENT FILING DATE: 2003-04-09
NUMBER OF SEQ ID NOS: 267
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 247
LENGTH: 9369
TYPE: DNA
ORGANISM: Herpes simplex virus
PCT-US03-11231-247

Query Match 3.9%; Score 48.8; DB 1; Length 9369;
Best Local Similarity 46.1%; Pred. No. 0.005;
Matches 201; Conservative 0; Mismatches 232; Indels 3; Gaps 1;

QY 313 CGGCTCGCAAGCTGCTCAACCGCTGTTAGTCAAGCGGCAATGACCTGCTGCGCGC 372
DB 4489 CGCGATGCGCTGAGAGCGCATCCCGGCGGAGTGCACCGGCGCGCTGCGCGCGCTG 4548
QY 373 GAAATACAGCGACCGTGAACGCTGCTGATGCTGCTGCGGAGCAAGAGAGTTGAC 432
DB 4549 GACCTGGCGCGCGCGCGGAGATATCCGCGCGCGGAGATGACCGGAGCGGCGG 4608
QY 433 GTTGTGCGGAGATTACCGGAGGAGAAATCCGATGCGCGGAGATGACGCTGTTGAAGTT 492
DB 4609 GAGCTGTGTCGAGACCGGAGGTACCTGCTGTGCTCCAGAGAGAGCGCTGTGCTGCTG 4668
QY 433 CGGCGGATGATGAGAGAAATTCGCTGCTTCCGCTGCGGAGATGACCGCGCTGCGG 552
DB 4669 GCGGCGGCTTGAAGACCGGATGACGAGCCCGGAGAAACAGCGCTGACGAGATAT 4728
QY 553 GTGCTTTGTCGCGC---CAGTTCATGAGAGAGACCAAGACCTGCTGCTGCTGAC 609
DB 4729 GTGCTTTCATGACCGCGGAGAGACGCGCGGAGATCAAGAGAGAGAGTGTGCGCGCAA 4788


```
OY 610 GAGGGCTGCGCTCTCCATGACGCTCTCATGAGCGCGCAGAACCCGCTCGAAAT 669
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4789 CAGCAGCCGCGGAGCCACGAGCGGGTCATGGCGGCTTGCGGAGCGCTGGCCGG 4848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 670 GACCTTGAAGATCTCTTACGCGCCGAGCGCGAGAGGCTGAGCAGAGAG 729
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4849 CGCGAGCCCGGCGCCCAATGAGCGCGAGCGCTGGCCMACCTCAAGACATGCTGAAG 4908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 730 CTGCTCGCGCTCTGATG 745
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4909 GTGCTCGCTCCGCG 4924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15

US-10-425-114A-14511

; Sequence 14511, Application US/10425114A

; GENERAL INFORMATION:

; APPLICANT: Liu, Jindong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114A

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 14511

; LENGTH: 1036

; TYPE: DNA

; ORGANISM: *Chlorella vulgaris*

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB191-016-B10_FLI

US-10-425-114A-14511

Query Match 3.8%; Score 48; DB 6; Length 1036;

Best Local Similarity 48.9%; Pred. No. 0.0041;

Matches 129; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

```
OY 448 GCGAGGAATCCGATGCGCGCATGACGCGCTCTGTGAAAGTTCCGCGCGAGTGTGAC 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 116 GGGGAGAGTTGGCATATTGCCACCTTCGCGCCCAAGCCCATGAGGGCGACTGGAAC 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 508 GAGAAATTCCGTGCTTGGCTCGGCGACTGCGCGCGCGCTGGCGCTTGGTGGCCC 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 176 GGCACCGCGCGCACCACTACTTCACGAGAGCATGCGCAAGGACCGCGCATGGCT 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 568 CAGTCAATGAGAGACCAAGACCTGTGCGCTCCGTCAACGAGGGGCTCGCGTGTCTC 627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 236 GCGATTGAGAGACGCAATGAGAGCTGTCCAAAGTCCACCGAGACATTTGCGCATAC 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 628 CATGACGTCTCTGATGAGCGCGCGCAGAACCCGCTCGAAATGACGTCTTGAAGATGCTG 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 296 GGGTACGGGAACGAGAGGGGCTGACCGGCAAGCAGAGCATGCGACATCAACACTTC 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 688 CTTCAAGCCGAGCGCGCAGCGCAGC 711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 356 CGCTACGGGCTGGCCGACCGCGGC 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: October 4, 2003, 12:01:00
Job time : 233.454 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: October 4, 2003, 03:31:13 ; Search time 289.856 Seconds
(without alignments)
11088.586 Million cell updates/sec

Title: US-10-014-717-1_COPY_62369_63628

Perfect score: 1260
Sequence: 1 atgcacagagagcaagcga.....agccctcaagctcgatag 1260

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09C_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	100.0	68750	13	US-10-014-717-1 Sequence 1, Appl1
2	151.4	12.0	1227	12	US-10-214-446-3 Sequence 3, Appl1
3	149.4	11.9	1233	14	US-10-205-032-7 Sequence 7, Appl1
4	149.4	11.9	60196	14	US-10-205-032-1 Sequence 1, Appl1
5	148.8	11.8	1164	14	US-10-156-761-6226 Sequence 6226, Ap
6	148.8	11.8	9025608	14	US-10-156-761-1 Sequence 1, Appl1
7	146.4	11.6	1197	14	US-10-156-761-409 Sequence 409, Appl
8	146.4	11.6	100000	14	US-10-156-761-15103 Sequence 15103, A
9	146.4	11.6	9025608	14	US-10-156-761-1 Sequence 1, Appl1
10	142.6	11.3	1215	12	US-10-214-446-49 Sequence 49, Appl
11	135	10.7	1248	12	US-10-214-446-55 Sequence 55, Appl
12	134.8	10.7	1248	12	US-10-214-446-55 Sequence 45, Appl
13	132	10.5	1197	14	US-10-156-761-2364 Sequence 2364, Ap
14	131.2	10.4	1179	12	US-10-214-446-31 Sequence 31, Appl
15	129.4	10.3	1278	12	US-10-214-446-19 Sequence 19, Appl
16	127	10.1	1179	14	US-10-156-761-1975 Sequence 1975, Ap

17	123.4	9.8	1248	14	US-10-156-761-7447 Sequence 7447, Ap
18	123	9.8	1251	10	US-09-861-289-38 Sequence 38, Appl
19	123	9.8	1251	10	US-09-860-846-38 Sequence 38, Appl
20	123	9.8	1251	11	US-09-988-384B-38 Sequence 38, Appl
21	123	9.8	1251	11	US-09-836-821-38 Sequence 38, Appl
22	123	9.8	5970	12	US-10-793-708-21 Sequence 21, Appl
23	123	9.8	5970	12	US-10-201-365-11 Sequence 11, Appl
24	120.4	9.6	5970	12	US-10-160-539-21 Sequence 21, Appl
25	120.4	9.6	1191	14	US-10-156-761-7165 Sequence 7165, Ap
26	120	9.5	1215	12	US-10-214-446-39 Sequence 39, Appl
27	119.8	9.5	1209	14	US-10-205-032-9 Sequence 9, Appl1
28	119.8	9.5	60196	14	US-10-205-032-1 Sequence 1, Appl1
29	119.6	9.5	1215	14	US-10-156-761-4523 Sequence 4523, Ap
30	118	9.4	1212	14	US-10-132-134-22 Sequence 2881, Ap
31	118	9.4	125746	14	US-10-156-761-15102 Sequence 15102, A
32	116	9.2	1221	12	US-10-214-446-1 Sequence 1, Appl1
33	115	9.1	987	14	US-10-132-134-24 Sequence 24, Appl
34	115	9.1	1248	14	US-10-156-761-408 Sequence 408, App
35	115	9.1	1700	12	US-10-132-134-22 Sequence 22, Appl
36	114.6	9.1	1428	14	US-10-145-415-15 Sequence 15, Appl
37	114.4	9.1	1233	12	US-09-953-348-24 Sequence 24, Appl
38	114.4	9.1	1233	14	US-10-267-255-24 Sequence 24, Appl
39	114.4	9.1	12249	12	US-09-953-348-74 Sequence 74, Appl
40	114.4	9.1	12249	14	US-10-267-255-74 Sequence 74, Appl
41	114.4	9.1	18331	12	US-09-953-348-96 Sequence 96, Appl
42	114.4	9.1	18331	14	US-10-267-255-96 Sequence 96, Appl
43	111.8	8.9	1293	14	US-10-145-415-7 Sequence 7, Appl1
44	111.8	8.9	1293	14	US-10-145-415-19 Sequence 19, Appl
45	111.4	8.8	1428	14	US-10-145-415-21 Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Liagon, James
APPLICANT: Molnar, Istvan
APPLICANT: Finkle, Rose
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/10/014,717
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US/09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
ORGANISM: Sorangium cellulosum
US-10-014-717-1
Query Match 100.0%; Score 1260; DB 13; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCACAGAGCAAGCAAGCAATCAAGTGAACGCAAGCAAGCTGTTTCGACTTCAACCGCTTC 60
DB 62369 ATGCACAGAGCAAGCAAGCAATCAAGTGAACGCAAGCTGTTTCGACTTCAACCGCTTC 62428
QY 61 GCGCTGGTACGCGGAGGAGCAGCTTCCCGCGATGAGCGCTGAGAGAGCAACCCCC 120
DB 62429 GCGCTGGTACGCGGAGGAGCAGCTTCCCGCGATGAGCGCTGAGAGAGCAACCCCC 62488
QY 121 ATCTTACTGGAGTGAAGGCGGCTCTGAGTCTTCAACCGATNACGAGCTGCGGC 180

Db 62489 ATCTTCTACTGGAGATGAAGGCGGCTCTGAGGTCCTACCCGATACCCAGACGTCGCGC 62548
Qy 181 GTGTTCCGCGACGAACGCTTTCGCGGTCACTCGAAGAAAGTGGGAATCGAGCGCGAGTAC 240
Db 62549 GTGTTCCGCGACGAACGCTTTCGCGGTCACTCGAAGAAAGTGGGAATCGAGCGCGAGTAC 62608
Qy 241 TGGTGGCCATTCGCGAGCTCGAGCTATGAAGAAAGTGGGAATCGAGCGCGAGTAC 300
Db 62609 TGGTGGCCATTCGCGAGCTCGAGCTATGAAGAAAGTGGGAATCGAGCGCGAGTAC 62668
Qy 301 GAGGATCAAGCTCGAGCTCGAGCTATGAAGAAAGTGGGAATCGAGCGCGAGTAC 360
Db 62669 GAGGATCAAGCTCGAGCTCGAGCTATGAAGAAAGTGGGAATCGAGCGCGAGTAC 62728
Qy 361 CTGCTGCGCGCGAATAACAGCGCACCGTGCACAGCTCTCGATGCTCTCGCGACAA 420
Db 62729 CTGCTGCGCGCGAATAACAGCGCACCGTGCACAGCTCTCGATGCTCTCGCGACAA 62788
Qy 421 GAGGATCAAGCTCGAGCTCGAGCTATGAAGAAAGTGGGAATCGAGCGCGAGTAC 480
Db 62789 GAGGATCAAGCTCGAGCTCGAGCTATGAAGAAAGTGGGAATCGAGCGCGAGTAC 62848
Qy 481 CTGTTGAAGGTTCCGCGCGAGTGCAGAGAAATTCTCGCTCGGCTCGGCGACTGCG 540
Db 62849 CTGTTGAAGGTTCCGCGCGAGTGCAGAGAAATTCTCGCTCGGCTCGGCGACTGCG 62908
Qy 541 CGCGCGCTCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 600
Db 62909 CGCGCGCTCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 62968
Qy 601 TCCGTCACGAGAGGCGTCCGCGCTGCACAGTACGTCGTCGATGAGAGGCGCGAGAAACCG 660
Db 62969 TCCGTCACGAGAGGCGTCCGCGCTGCACAGTACGTCGTCGATGAGAGGCGCGAGAAACCG 63028
Qy 661 CTGCAAAATGAGCTCTTGAAGTGTCTTCAAGCGCGAGCGAGCGAGCGAGCTGAGC 720
Db 63029 CTGCAAAATGAGCTCTTGAAGTGTCTTCAAGCGCGAGCGAGCGAGCGAGCTGAGC 63088
Qy 721 ACGAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
Db 63089 ACGAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63148
Qy 781 TACCTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 63149 TACCTTATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63208
Qy 841 AAGGCGGAGCGCGGCTCATAGAGAAAGCGCTCGATGAGAGTGTCTCGCTTCGACATATC 900
Db 63209 AAGGCGGAGCGCGGCTCATAGAGAAAGCGCTCGATGAGAGTGTCTCGCTTCGACATATC 63268
Qy 901 CTCAGAAATGGAATCTGTGCTTTCGCGAGCGAGACCTGGAATGCTGCGCGGATGATC 960
Db 63269 CTCAGAAATGGAATCTGTGCTTTCGCGAGCGAGACCTGGAATGCTGCGCGGATGATC 63328
Qy 961 AAGAAAGGAGAGTGTCTTTCCTGATCCGAGGCGCTGAGAGTGGGACTGATATC 1020
Db 63329 AAGAAAGGAGAGTGTCTTTCCTGATCCGAGGCGCTGAGAGTGGGACTGATATC 63388
Qy 1021 TCCAGGCGCAGACGCTGTTTATGATGCGAGGAGACCGGCGCGAGCTCGCTGATAG 1080
Db 63389 TCCAGGCGCAGACGCTGTTTATGATGCGAGGAGACCGGCGCGAGCTCGCTGATAG 63448
Qy 1081 GGGCCCCCATGTCTGCGCGCGGAGTGTCTTGTCTGCTGCGCTGAGCGGAGATGCGCGTGGCG 1140
Db 63449 GGGCCCCCATGTCTGCGCGCGGAGTGTCTTGTCTGCTGCGCTGAGCGGAGATGCGCGTGGCG 63508
Qy 1141 ACCATCTCGATGCTTCCCGAGATGAAGCTGAAGAAATCCGCTGTTTGGATACAC 1200
Db 63509 ACCATCTCGATGCTTCCCGAGATGAAGCTGAAGAAATCCGCTGTTTGGATACAC 63568
Qy 1201 CCCGCTTCGGAACATCGAATCACTCAAGCTATCTTGAAGCGCTCAAGAGCTGATAG 1260
Db 63569 CCCGCTTCGGAACATCGAATCACTCAAGCTATCTTGAAGCGCTCAAGAGCTGATAG 63628

RESULT 2
US-10-214-446-3
; Sequence 3, Application US/10214446
; Publication No. US20030180742A1
GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 3
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Bacterial
US-10-214-446-3
Query: Match 12.0%; Score 151.4; DB 12; Length 1227;
Best Local Similarity 49.9%; Pred. No. 5.7e-34;
Matches 513; Conservative 0; Mismatches 486; Indels 30; Gaps 4;
Qy 147 CTGGTCTCTACCCGATACACAGACGTGTGCGCGGTGTTCCGCGACGAACGCTTGGCGGT 206
Db 138 CTGGTCTCTACCCGATACACAGACGTGTGCGCGGTGTTCCGCGACGAACGCTTGGCGGT 197
Qy 207 CAGTGAAGAGAGTGGAGATGAGCGCGAGAGTCTGTCGCGCATTCGCGAGCTCAGCGA 266
Db 198 CAGCGCGGTGCGCGTCAACGAGCTGAGAGAGAGAGTGGAGAGAGTGGAGAGAGTGGAGAG 254
Qy 267 TATGAAGAGAGTGGAGATGAGCGCGAGAGTCTGTCGCGCATTCGCGAGCTCAGCGA 326
Db 255 CGTCTGATGAGCAGACATGCTGTCGCGCAGCGAGAGACACACCGGCTGCGACCT 314
Qy 327 CGTCAACCGCTGTTTACGTACGCGCCATGACCTGTCGCGCGCGGAAATPACGCGAC 386
Db 315 CGTCTGAAAGGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 374
Qy 387 CGTCAACCGCTGTCGATGCTGCTGCGCGACAGAGAGTTCGAGCTGTGCGGATTA 446
Db 375 CACCGACCGCTTCTGCGAGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
Qy 447 CGCGAGGAGATCCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 506
Db 435 CG 494
Qy 507 CGAAGAGTTCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566
Db 495 GCCCGCATGCG 536
Qy 567 CCAGGTCGATGAGAGACCAAGACCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 626
Db 537 GAAACACCG 596
Qy 627 CCATGACGCTCTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686
Db 597 CGAGCGGTTGATGCG 650
Qy 687 GCTTACG 746
Db 651 GCTTTCG 710
Qy 747 TCGATTTATGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806

Db 711 CTTGCTACGCTCGCGGCGACGAGACGCGCCGCTGATCGCAACGCGGTGACGCG 770
Qy 807 CTTGCTACGCTCGCGGCGACGAGACGCGCCGCTGATCGCAACGCGGTGACGCG 866
Db 771 CTTGCTACGCTCGCGGCGACGAGACGCGCCGCTGATCGCAACGCGGTGACGCG 830
Qy 867 CGCGCTGATGAGGTGCTCGCTGATGAGACATCTCTGAAATAGAACTGTGCTTTCG 926
Db 831 CGCGATGAGAACTCTGCGCTGAGAGGCGCGGTGAGCGGCGGTGCGCGCTTAC 890
Qy 927 CAGGACGACCTGAGTACTGCGGCGCATTCATCAAGAAAGGAGATGTCTTCTCT 986
Db 891 CACGACCCGTAAGATGCGGCGGTGACCGTACCGCGCGGAGATGATCATCGG 950
Qy 987 GATCCGAGCGCGCTGAGATGAGATGAGTATTCCTGACGCGGAGCGTGTGA---TGT 1043
Db 951 GCTGCGCGCGCAATTCGAGACCGCGCGCTGACGACCGTCCGACATCTGAGATTCG 1010
Qy 1044 GCGACGAGACAGCGGCGCGAGCGCTGCGTACGCTAGAGGCGCGCATGTCTGCGCGGAGT 1103
Db 1011 ACGCGTGAAGTGGCGGCAACGCTGCTTTCGCGCATGGCGTGCATCTTCTGCGGTGC 1070
Qy 1104 GTCCCTTCTGCTGCTGAGGCGGAGATGCGGTGCGGACCATCTTCCGTAGTTCGCGA 1163
Db 1071 GCGCGTGGCGCGCGCGAGCGCGCGGATCGCATCGGACCGCTGCTGCGCGCTTCCCGA 1130
Qy 1164 GATGAGCT 1172
Db 1131 TCTGGGCT 1139

RESULT 3
US-10-205-032-7
; Sequence 7, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Stafia, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-205
; CURRENT APPLICATION NUMBER: US/10/205, 032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1233
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-7

Query Match 11.9%; Score 149.4; DB 14; Length 1233;
Best Local Similarity 49.7%; Pred. No. 2.2e-33;
Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;
Qy 288 CGGGGTCGCGCGGAGATCAAGCTCGGCTCGCAACCGCTGTTAGTC 347
Db 261 CGGATGACCGCGCGGCGCGCGCGCTGCGAGCTGAGAGCGGCTTACCCC 320
Qy 348 ACGGCGCATGACCTGTGCGCGCGCGCAATACAGCGACCGTGCAGCTGCTGATGC 407
Db 321 CGGCGGATGCGCGCGAGTGAAGTGCAGTGGGAGATCAACCGCGCGAGTGTGACCA 380
Qy 408 TCGCTCGGACAAAGAGATTCAGCTTTCGCGGATTAACGCGAGGAAATCCGATCG 467
Db 381 GGTACGACACGAGACCGGATGACATGCGACGACCTCGGTACCGCGTGCAGTAC 440
Qy 468 CGCGATGACGCTGTTGAAGTTTCGCGCGAGTGTACAGAGAAATTCGCTTCG 527
Db 441 GGTATGCGCGAGCTGCTGCGCATTCACCAAGAGATCAAGAAATTCGCGAGTGGT 500
Qy 528 CTCGCGACCTGCGCGCGCTGCGGTGGTGGTGGTCCGAGTGCATGAGAGACCA 587

Db 501 GACATCATCTTCAAGCAACGAGGCTGAGATATCCAACTCCCGGACGATTCACCGA 560
Qy 588 GACCTGTGCTGCTGCTGCAACCGAGGCGCTGCGGCTGCTCATGATGATGACG 647
Db 561 GACGTGCGCGCGCGCATGAGAGTGTGCGAAATCTTGAACCGAGATGCGCACAA 620
Qy 648 GCGAGAGAACCGCTGCAAAATGACGTCTTACAGATGCTGCTTACAGCGCGAGCG 707
Db 621 GCGCGCGGACCG---AAGAGACATGATAGAGGCGCTTGTGCGGCGAGTGCAGG 677
Qy 708 CAGAGGCTGAGCAAGAGAGTGTGCGCTGCTGCGGTGCGATTAATGCTGTCGAC 767
Db 678 GCGAGAGCTGACCGACGAGAGATGCTACATGTCGCGCTGCTCACCGCGGCGCA 737
Qy 768 CGATACACAGATTAACCTTAAGCGGTTCGCTGCTCAACCTGCTGCGGTCCGAGG 827
Db 738 CATTCGACCGCGCGCTGCTGCAACCTGTTCTGCTGCTGAGAGACCGCGAGG 797
Qy 828 GCTGAGCTGAGAGCGCGCGCGCGCTGATGAGAACCGCTGATGAGTGTGCG 887
Db 798 AAGAGCGCGGTGCGCGCGCGCGCGCGCGCTGCTGCTGAGAGAGAGTGTGCG 857
Qy 888 CTTCGACATATCTTCAAGATGAGAACTGTGCTTTCGCGAGCGAGAGTACTG 947
Db 858 CT---ACCGGTCCCGTTCAACTGATCTTCCGATCTTGAACGAGACACCGATCT 914
Qy 948 CGGGCATCATCAAGAAAGGAGATGTCTTCTCTGATCCCGAGCGCTGAGAGA 1007
Db 915 CGGCGACCCCATGCGCAAGGCGCAGATGTGATGCTCTGATGCTTCCGGAACCGG 974
Qy 1008 TGGAGTGTATTTCTCAAGCGCGAGAGTGTGATGAGCGGACAGCGCGAGCT 1067
Db 975 CACGAGGTGTTCAAGACCGGACACCTTGCATCGAGCGAGTGAACAGACCT 1034
Qy 1068 CGCGTACGTAGAGCGCGCGCGCGCGCGCGCGGTGCTTCTGCTGCGCTGAGGCG 1127
Db 1035 GCGCTGCGCGACGCGCATTCACGACTGCTGCGCGGTCTTGCAGCGAGTGAAGCG 1094
Qy 1128 GATGCGCGTGGGACCATCTTCCGTAGTTCGCCGAGAT 1166
Db 1095 GGTCTTCTCAACGAGCGCTGCGACGATTCACGAGTT 1133

RESULT 4
US-10-205-032-1/C
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Stafia, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-205
; CURRENT APPLICATION NUMBER: US/10/205, 032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-1

Query Match 11.9%; Score 149.4; DB 14; Length 60196;
Best Local Similarity 49.7%; Pred. No. 4.3e-33;
Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;
Qy 288 CGGGCTCGCGCGGAGATCAAGCTCGGCTCGCAACCGCTGTTAGTC 347
Db 3601 CGGATGACCGCGCGCGCGCGCGCTGCGAGCTGAGAGCGGCTTACCCC 3542
Qy 348 ACGGCGCATGACCTGTGCGCGCGCGGAAATACGCGACCGTGCAGAGCTTGCATGC 407

```

Db 3541 CCGCCGATCGCCAGATGAGTGGATCGGCGATGACCGGATCGTCTGACCA 3482
Qy 408 TCGCTCCGACAAAGAGAGTTCAGATTCGCGGATTAACGGAGGAATCCGATGCG 467
Db 3481 GGTACGCGACCAAGACCGGATGACATCGCAGCGACCTCGGTAACCGCTCCGATGAC 3422
Qy 468 CGCGATCAAGCTCTGTTGAAGTTCCGCGGATGACGAGAAATTCCTGCTTCG 527
Db 3421 GGTCAATCGCGAGCTGCTCGGCAATCCACAAAGATCAAGAAATTCGCGAGTGGT 3362
Qy 528 CTGCGGATGCGCGCGCGCTGCGCGGATTCGCGGATTCGCGGATTCGCGGATTCG 587
Db 3361 GGCATCATCTCAAGCAAGAGGCTGAGATCCCAACTCCGCGAGATCTTACCGA 3302
Qy 588 GACCTGTGCGGCTCGCTGACCGAGGAGCTGCGCTGCTCATGAGCTTCGATGAGCG 647
Db 3301 GACGCTGCGCGCGCGCTGAGAGATGTCGAAATTCCTGTAAGCCAGATCGCCACAA 3242
Qy 648 GCGCAGAAACCGCTCGAAATGAGCTCTTGAAGATGCTGCTTCAAGGCGAGCGAGCG 707
Db 3241 GCGCGCGGAAACG---AAGAGACCTGATCAAGCGGCTCTGTCGCGCGAGAGTGCAGCG 3185
Qy 708 CAGCAGCTGAGACGAGAGAGTGTGCGGCTGCGGATTCGCGGATTCGCTGCGAGC 767
Db 3184 GCGCAAGCTGACGAGAGAGAGTGTCAACATCTGTCGCTGCTGCTGCTGCTGCTGCTG 3125
Qy 768 CGATACCAAGATCTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 827
Db 3124 CATCTCAAGCTGACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3065
Qy 828 GCTCAGCTGCTGAGAGCGGAGCGGCTCATGAGAAACGCGCTTGAAGATGCTGCG 887
Db 3064 AAGGCGCGGCTGCGCGCGGAGCGAGCTGCTGCGGAGCTGCTGCTGCTGCTGCTGCTG 3005
Qy 888 CTTCGCAATATCTCAAGATAGAACTGTGCTTTCGCGAGAGAGCTGAGATGCTG 947
Db 3004 CT---ACCGGCTCCGCTTCACTGATCTTCGCGATCTGAGAGAGACCAAGATCTCT 2948
Qy 948 CGGCGCATGATCAAGAAAGGAGATGCTTCTCTGATTCGCGAGCGCTGAGAGA 1007
Db 2947 CGGCGCATGATCAAGAAAGGAGATGCTTCTCTGATTCGCGAGCGCTGAGAGA 2888
Qy 1008 TGGAGCTGATTCCTCAAGCGAGCTGTTGATGTCGAGAGAGAGAGAGAGAGAGAG 1067
Db 2887 CACCGAGGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2828
Qy 1068 CGCGTACGAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1127
Db 2827 GCGCTTCCGCGACGAGATCAACATCTGCTGCGAGAGAGAGAGAGAGAGAGAGAG 2768
Qy 1128 GATCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1166
Db 2767 GGTCTTCTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2729

```

RESULT 5
US-10-156-761-6226

; Sequence 6226, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

```

; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6226
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1164)
US-10-156-761-6226

```

Query Match 11.8%; Score 148.8; DB 14; Length 1164;

Best Local Similarity 50.0%; Pred. No. 3.2e-33;

Matches 449; Conservative 0; Mismatches 422; Indels 27; Gaps 2;

```

Qy 284 TGTTCGAGCTGCGCGAGAGATCAAGCTGAGTCCGCAAGCTGCTCAACCGTCTGTTA 343
Db 223 TGTTCATATGAGAGCGCGCGAGACCAACCGCATCTGCTGCTGCGCGGCTTCA 292
Qy 344 CGTACGCGGATGACCTGCTGCGCGGAAATACAGCGACCGTGCACGACTGCTCG 403
Db 293 CCTCTGCTGCGGATGACCAATCCGCGAGCCGCTACGAGAGACCGCCACCGGCTTTCG 352
Qy 404 ATGCTGCTCGCGACCAAGAGAGATTCGATGCTGCGGAGATTCGCGGAGAGAAATCCGA 463
Db 353 ACGCATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412
Qy 464 TCGCGGATGACGCTGCTGTTGAAGTTCGCGCGAGAGTGCAGAGAGTTCGCTGCT 523
Db 413 TCACCTGATCTGCACTGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 472
Qy 524 TCGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 583
Db 473 GAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 508
Qy 584 CCAAGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 643
Db 509 CGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 568
Qy 644 AGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 703
Db 569 ACAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 625
Qy 704 ACGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 763
Db 626 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 685
Qy 764 GCACGATACAGAGATCACTTATCGGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTG 823
Db 686 GCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 745
Qy 824 AGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 883
Db 746 AGCACTGCGCGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805
Qy 884 TCGCTTGAAGATATCTCAAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 943
Db 806 CCGCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 865
Qy 944 ACTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1003
Db 866 TCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 925
Qy 1004 GAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1063
Db 926 GTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 985
Qy 1064 GCTTCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1123
Db 986 ACTCGCGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1045

```


Qy 1124 CGAGATCGCCGCGGAGCAACATCTTCGTAAGTTCCCGAGATGAAGCTGAAGAAC 1181
Db 1046 CGAGATCGCCCTCGCGCGCTCTCGAACGATTCGCCGAGACTCGCCTCGCGGAAC 1103

RESULT 6
US-10-156-761-1

; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 11.8%; Score 148.8; DB 14; Length 9025608;
Best Local Similarity 50.0%; Pred. No. 1.5e-32;
Matches 449; Conservative 0; Mismatches 422; Indels 27; Gaps 2;

Qy 284 TGTTCGGGCTCCCGCGGAGATCAGCTGGGCTCCGAACTGCTCAACCCGTCGTTA 343
Db 7508344 TGCTCAACATGAGACGCGCGGACACACCCGATCCGCGCTGTCGCGCGGCTTCA 7508403
Qy 344 CGTACGCGCCATGACCTGTGCGCGCGCGAAATTAACAAGCAAGCTGACAGCTGCTG 403
Db 7508404 CCTGCGCTCGGTGAGAACATCCGCGAGCCCGTACGAGAGACGCCACCGGCTTCTG 7508463
Qy 404 ATGCTCGCTCCGGAACAAGAGATTGACGTTGTGCGGATTACGCGAGGAAATCCCGA 463
Db 7508464 ACGCACTCGGAGACGAGCGGACGACCTCATGCGCTCGTAAGCCGCGCTGCGA 7508523
Qy 464 TCCCGCGATAGCGGCTGTGTAAGTTCCGCGCGAGTGTGACGAGAAATTCGTCCT 523
Db 7508524 TCACCTGATCTGCGACCTGCTCGGCGTCCGGAACAACCGGCGGACTTCCGAGCT 7508583
Qy 524 TCGGCTCGGAGCTGCGCGCGCGCTCGGCGTGTGTTGTCGCCAGTGCATAGAGAGA 583
Db 7508584 GGACCGACCCGC-----TGTCAACCCCGAGACCCGCGCCGC 7508619
Qy 584 CCAAGACCTGTCGTCGCTGCTACCGAGGAGCTGCGCTCTCCATAGCTCTGCATG 643
Db 7508620 CGGACGTGCGAGGAATTCGTCGTGTGTTGCTTGCTTCAACGCGGCTCTCGCG 7508679
Qy 644 AGCGCGGAGAACCCGCTGGAATAAGCTGTGACGATGTGCTTACGCGCGAGCGG 703
Db 7508680 ACAAGCGGAAACCC---CGCGAGACCTGCTCTCGACCTCATCCCGTGAAGAG 7508736
Qy 704 ACGGAGAGGCTGAGCAAGAAGAGTGTGCGGCTGTGGGTGCGATTATGCGTGTG 763
Db 7508737 AGGGGAGCGGCTCACGAGAGCGAGTGAATGCTCTGCTTCTCATCTTTCGCG 7508796
Qy 764 GCACGATACCAAGATTAATCTTATGCGTTCGCTGTCTCAACTGCTGCGGTCGCG 823

Db 7508797 GCTACGAGAACACCTGTCACCTCATCGGAACCGCGTAATGCGCCTGTCGCCATCCG 7508856
Qy 824 AGCGCTCGAGCTGTGTAAGCCGAGCCCGGCTCATGAGAAACGCGCTGATAGTGC 883
Db 7508857 AGCAGCTCGCCGCTCTCGGAGAGACCGGACCGCTGCCAGACCGCTCGGAGTTG 7508916
Qy 884 TCCGCTTCGACATATATCTTGAATAGAACTGTGCTTTGCGCAGCAGGACCTGAGT 943
Db 7508917 CCGCTACGAAGAGCCCGCTGCTGTCATCCGCGCTTCGCGTGGGACGTGACGA 7508976
Qy 944 ACTGCGGGGATCGATCAAGAAAGGAGATGTTCTTCTCTATCCGAGCGCTCGA 1003
Db 7508977 TCGCGGGGTCAACGTCGCCGCGGGAGACCGGTGCTGTCTCTGTCGCCCAAC 7509036
Qy 1004 GAGATGGACGTATTTCTCAAGCGACAGCTGTGATGTGCGACGGGACACGGGCGGA 1063
Db 7509037 GTGACCGAGCGGCTTCCGACCCGACCGCTGTGATCTGCGCGGACGCGCGAGCC 7509096
Qy 1064 GCTCGCGTACGATGAGAGCCCGCATGTGCCCCGCGGTGTCCTTGTGCTGCTGAG 1123
Db 7509097 ACTTCGCGCTCGGCGACGCGTCACTGCTGCGCGCGCTGCGCTGCGGCTGAGA 7509156
Qy 1124 CGAGATCGCCGCGGAGCAACATCTTCGTAAGTTCCCGAGATGAAGCTGAAGAAC 1181
Db 7509157 CGAGATCGCCCTCGCGCGCTCTCGAACGATTCGCGAGCTCGCCTCGCGGAAC 7509214

RESULT 7
US-10-156-761-409

; Sequence 409, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 409
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1197)
US-10-156-761-409

Query Match 11.6%; Score 146.4; DB 14; Length 1197;
Best Local Similarity 50.8%; Pred. No. 1.6e-32;
Matches 453; Conservative 0; Mismatches 406; Indels 33; Gaps 3;

Qy 279 CGGATTTTGGGGTGGCGCGCGAGATCAAGCTCGGAGTCCGCAACCCGTC 338
Db 255 CAGCTGTGTGAGAGAGAGAGCGGAGACACTCGAGCGAGCTGCTGCGAAGA 314
Qy 339 GTTACGTCACGCGCATGACCTGTGCGCGCGAAATACAGCGCACCGTGCACGACT 398
Db 315 GTTACGTCAGCGGAGAGAGAGCGCTGCGCGCGAATCATCAGGCAATCGTCAGAGCA 374
Qy 399 GCTGATGCTGCTCCGAGCAAGAG--AGTTGACGTTGTGCGGAGTTACGCGGAGG 455
Db 375 CTTGATGTCATCGAGCGCGGCGCGCGTGTGACCTGTGTAAGACTTTCGCAAGCG 434

```
QY 456 AATCCGATCCGCGCATGAGGCTTGTGAAGTTCCGCGCGATGTGACGAAAGTT 515
DB 435 GGTGCGCTCATGTGATTTCCGACCTGTCGATGCCGTCGACGCGCGGAGTT 494
QY 516 CCGTCCCTTCCGCTCGGCTGAGCTGCGCGCGCTCGGCTGGGTTTGGTCCCGAGTCA 575
DB 495 CAGACATCCGCGGCGCATGATGCG-----GGTCA 527
QY 576 TGAAGAGACCAAGACCTGTCGCTGCTCACCGAGGGCTCGCGCTCCATGACGT 635
DB 528 CAGAGCGCGCGCGACCGAGGCGCGCGCATGCGCTGGCGGCTCTCTACAGAGCT 587
QY 636 CCGTCAATGAGCGCGCGAGAAACCGCTCGAAAATGACGTTGACATCTCTTCAAGC 695
DB 588 TGTTCAGAGCGCGCGCGCAACCC--CGGTGACGCTGATCTCGGCTGATACACAC 644
QY 696 CAGAGCGCGCGCGAGGCTGAGCAGAAAGAGTGTGCGGCTCGTGGTGGCATTA 755
DB 645 CAGAGACCCCGATGATGTGTGAGACATGTTTCTCATGAACGCGCGCGGACCTGCT 704
QY 756 CGCTGTCGACCGATACGATCTACCTTATCGGCTTCTGCTGCTCAACCTGTCG 815
DB 705 GATAGCGGCGCGACGACACCGCTGATGATCGGCTCGGTCGCGCTGCTCTCGA 764
QY 816 GTCCGCGCGCGCGCTGAGCTGTGAAGCCGAGCCGCGCTCATGAGAAACGCTGCA 875
DB 765 CAGCCCGCGACGAGCTGCGCTGTCGCGAGACCCGTCGTGTGCGCAACGCTGCA 824
QY 876 TGAAGTCTCCGCTTTCAGACATATCTCAGAAATAGAACTGTCGCTTCCGAGGAGA 935
DB 825 GAGAGCTGCGCTGCTACTACATCGGCGGAGCTTCCGCGCGCGAGCGCTGCGACCCGGA 884
QY 936 CCGTGAATCTGCGGCGCATTCGATCAAGAAAGGAGATGTTCTTCTGATCCGAG 995
DB 885 CCGTGAATCTGCGGCGCATTCGATCAAGAAAGGAGATGTTCTTCTGATCCGAG 944
QY 996 CCGCCCGAGAGATGGAATGTAATCTCCAGGCGAGCGTGTGATGTGACGCGGACAC 1055
DB 945 CCGCGACTTCGACCCCGCTTCTGAGAGAGCGGAACTTTCGACATCACTCGACGCC 1004
QY 1056 GAGCGCGAGCTTCGCTAGCGTGAAGCGCGCGCATGTCCCGCGGATGTCCCTTCTCG 1115
DB 1005 CCGCCCGCATGTGCTTGGCTTGGCGGACACGATGATCGGCGACCAACTGCGCG 1064
QY 1116 CCGTGAAGCGGAGATCGCTGCGGACCATTTCCGTAAGTTCCCGAATG 1167
DB 1065 GATCGAGCTGAGATGCTTCGAGACCTGTTCCGCGCTCCGCGGCTCG 1116
```

```
RESULT 8
US-10-156-761-15103
; Sequence 15103, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15103
; LENGTH: 100000
```

```
TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15103
Query Match 11.6%; Score 146.4; DB 14; Length 100000;
Best Local Similarity 50.8%; Pred. No. 3.5e-32;
Matches 453; Conservative 0; Mismatches 406; Indels 33; Gaps 3;
QY 279 CGGATTTGTCGAGCTGCGCGCGGAGATCAGCTCGGTCCGCAAGCTCTCAACCCGTC 338
DB 80684 CAGCTGTTGTGCGAGACGAGCCGAGACACCTCGGACCGCAAGCTCTGCGCAAGA 80743
QY 339 GTTACGTACGCGCGCATCGACTTGTGCGCGCGGAAATACGCGACCGTCAACAGCT 398
DB 80744 GTTACCGTACGCGCGCATCGAGCGCTGCGCGCGGAACTCAAGCGCATCTGAGAGCA 80803
QY 399 GCTCATGCTGCTTCGCGCAAGAG--AGTTGACGTTGTGCGGATTAACGCGAGGG 455
DB 80804 CCGTATGATCGATCGAGCGCGCGCGCGCGGCTGACCTGGTGAAGAC
```


QY 361 CTGCTGCGCGCGAATAACAGCGCACCGTCGACCACTGCTGATCTGCTCCGACAA 420
DB 62729 CTGCTGCGCGCGCGAATAACAGCGCACCGTCGACCACTGCTGATCTGCTCCGACAA 62788
QY 421 GAGGAGTTGAGGTTGCTGGGGGATTACGGGAGGGAATCCGATGCGCGGATCAGCGCT 480
DB 62789 GAGGAGTTGAGGTTGCTGGGGGATTACGGGAGGGAATCCGATGCGCGGATCAGCGCT 62848
QY 481 CTGTTGAAGTTTCCGCGCGAGTGTACGAGAAATTCCTGCTGCTGCTGCGGACTGCG 540
DB 62849 CTGTTGAAGTTTCCGCGCGAGTGTACGAGAAATTCCTGCTGCTGCTGCGGACTGCG 62908
QY 541 CCGCGCTGCGCGCTGGGTTTGGTGGCCCGAGTGTGATGAGAGAACCAACCTGCTGCGG 600
DB 62909 CCGCGCTGCGCGCTGGGTTTGGTGGCCCGAGTGTGATGAGAGAACCAACCTGCTGCGG 62968
QY 601 TCCGTCACGAGGAGGCTCGCGCTGCTCATGAGAGTCTGCTGATGAGGCGCGAGAACCCG 660
DB 62969 TCCGTCACGAGGAGGCTCGCGCTGCTCATGAGAGTCTGCTGATGAGGCGCGAGAACCCG 63028
QY 661 CTCGAAATGACGCTCTTGAAGTGTCTTCAAGCCGAGGCGCGAGCGGAGCTGAGC 720
DB 63029 CTCGAAATGACGCTCTTGAAGTGTCTTCAAGCCGAGGCGCGAGCGGAGCTGAGC 63088
QY 721 ACGAAGGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 63089 ACGAAGGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63148
QY 781 TACCTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 63149 TACCTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63208
QY 841 AAGGCGGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 63209 AAGGCGGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63268
QY 901 CTCGAAATGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 63269 CTCGAAATGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63328
QY 961 AAGAAAGGAGGAGTGTGCTTCTTCTGATCCGAGGCGCTGAGAGATGGAATCTGATTC 1020
DB 63329 AAGAAAGGAGGAGTGTGCTTCTTCTGATCCGAGGCGCTGAGAGATGGAATCTGATTC 63388
QY 1021 TCCAGGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB 63389 TCCAGGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63448
QY 1081 GCGCCCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 63449 GCGCCCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63508
QY 1141 ACCATCTTCCGATGCTTCCCGAGATGAGCTGAAGAAATCTCCGTTTGGATACAC 1200
DB 63509 ACCATCTTCCGATGCTTCCCGAGATGAGCTGAAGAAATCTCCGTTTGGATACAC 63568
QY 1201 CCGCGCTTCCGAGAACATCAATCACTCAACGTCATCTTGAAGCCTTCAAGCTGATAG 1260
DB 63569 CCGCGCTTCCGAGAACATCAATCACTCAACGTCATCTTGAAGCCTTCAAGCTGATAG 63628

RESULT 2
US-09-568-102-1

Sequence 1, Application US/09568102

Patent No. 6346404

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zirkle, Rose

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568.102
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-568-102-1
Query Match 100.0%; Score 1260; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9.9e-250;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACACAGGAGCAAGCGAATGAGTGAAGCAAGCGCTGCTTGAACCTTCAAGCGATTC 60
DB 62369 ATGACACAGGAGCAAGCGAATGAGTGAAGCAAGCGCTGCTTGAACCTTCAAGCGATTC 62428
QY 61 GCGCTGGGTAAGCGGAGGAGCCGTTCCCGGATGAGCGCGCTGAGAGGCAACCCG 120
DB 62429 GCGCTGGGTAAGCGGAGGAGCCGTTCCCGGATGAGCGCGCTGAGAGGCAACCCG 62488
QY 121 ATCTTACTGGAGTGAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 62489 ATCTTACTGGAGTGAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 62548
QY 181 GGTTCGCGAGCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
DB 62549 GGTTCGCGAGCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 62608
QY 241 TCGTGGGCAATCCGAGCTGAGCGATGAGAGATGAGATGAGATGAGATGAGATGAGATGAG 300
DB 62609 TCGTGGGCAATCCGAGCTGAGCGATGAGAGATGAGATGAGATGAGATGAGATGAGATGAG 62668
QY 301 GAGATCAAGCTGCGGCTGCGCAAGCTGCTCAACCCGTTGTTAGTCAACGCGCATTCAC 360
DB 62669 GAGATCAAGCTGCGGCTGCGCAAGCTGCTCAACCCGTTGTTAGTCAACGCGCATTCAC 62728
QY 361 CTGCGCGCGCGGAAATACAGCGCACGCTGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 62729 CTGCGCGCGCGGAAATACAGCGCACGCTGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 62788
QY 421 GAGGAGTTGAGAGTGTGCGGAGTTACGCGAGGGAATCCGATGCGCGGATCAGCGCT 480
DB 62789 GAGGAGTTGAGAGTGTGCGGAGTTACGCGAGGGAATCCGATGCGCGGATCAGCGCT 62848
QY 481 CTGTTGAAGTTTCCGCGCGAGTGTACGAGAAATTCCTGCTGCTGCTGCTGCTGCTGCTGCT 540
DB 62849 CTGTTGAAGTTTCCGCGCGAGTGTACGAGAAATTCCTGCTGCTGCTGCTGCTGCTGCTGCT 62908
QY 541 CCGCGCTGCGCGCTGGGTTTGGTGGCCCGAGTGTGATGAGAGAACCAACCTGCTGCGG 600
DB 62909 CCGCGCTGCGCGCTGGGTTTGGTGGCCCGAGTGTGATGAGAGAACCAACCTGCTGCGG 62968
QY 601 TCCGTCACGAGGAGGCTCGCGCTGCTCATGAGAGTCTGCTGATGAGGCGCGAGAACCCG 660
DB 62969 TCCGTCACGAGGAGGCTCGCGCTGCTCATGAGAGTCTGCTGATGAGGCGCGAGAACCCG 63028
QY 661 CTCGAAATGAGAGCTCTTGAAGTGTGCTTCAAGCGCGGAGCGGAGCGAGCGGATGAGC 720
DB 63029 CTCGAAATGAGAGCTCTTGAAGTGTGCTTCAAGCGCGGAGCGGAGCGGAGCGGATGAGC 63088
QY 721 ACGAAGGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 63089 ACGAAGGAGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63148
QY 781 TACCTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 63149 TACCTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63208

OY 841 AAGCCGAGCCCTGCTCATAGAGAACGCGCTCATAGAGTCTCCGCTTCACAAATATC 900
DB 63209 AAGCCGAGCCCTGCTCATAGAGAACGCGCTCATAGAGTCTCCGCTTCACAAATATC 63268
OY 901 CTCGAATAGGAACCTGTCCGCTTCCCGAGGACGAGCTGTAGTACTGCGGGGACATCCATC 960
DB 63269 CTCGAATAGGAACCTGTCCGCTTCCCGAGGACGAGCTGTAGTACTGCGGGGACATCCATC 63328
OY 961 AAGAAAGGGAGATGCTCTTCTCTCTGATCCCGAGCGCCCTGTAGAGATGGAGCTGATTC 1020
DB 63329 AAGAAAGGGAGATGCTCTTCTCTCTGATCCCGAGCGCCCTGTAGAGATGGAGCTGATTC 63388
OY 1021 TCCAGGCCAGACGTGTGTGATGTGACGGGACACGGGCGCGAGCTCGCGTACGGTATGA 1080
DB 63389 TCCAGGCCAGACGTGTGTGATGTGACGGGACACGGGCGCGAGCTCGCGTACGGTATGA 63448
OY 1081 GGGCCCCATGTCTGCCCCGGGGGTGTCCCTGTGCTCGAGGCGGAGATCGCCGTGGGC 1140
DB 63449 GGGCCCCATGTCTGCCCCGGGGGTGTCCCTGTGCTCGAGGCGGAGATCGCCGTGGGC 63508
OY 1141 ACCATCTTCGGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATATACAC 1200
DB 63509 ACCATCTTCGGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATATACAC 63568
OY 1201 CCCGCGTTCGGGAACATGAAATCACTCAACGTCACTTGAAGCCCTCCAAAGCTGGATAG 1260
DB 63569 CCCGCGTTCGGGAACATGAAATCACTCAACGTCACTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Lignon, James
APPLICANT: Molnar, Istvan
APPLICANT: Ziekler, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,969
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-567-969-1

Query Match 100.0%; Score 1260; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9,9e-290;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATGACACAGAGCAACGAATCAGTAGAGAGCAAGCTGCTTTCAGCTTCAAGCCGTTTC 60
DB 62369 ATGACACAGAGCAACGAATCAGTAGAGAGCAAGCTGCTTTCAGCTTCAAGCCGTTTC 62428
OY 61 GCGCGCTGGGTACGCGGAGAGACCCGTTCCCGGATCGAGCGCGCTGAGAGAGCAACCCCC 120
DB 62429 GCGCGCTGGGTACGCGGAGAGACCCGTTCCCGGATCGAGCGCGCTGAGAGAGCAACCCCC 62488
OY 121 ATCTTCTACTGAGTGAAGCGCTCTGTGCTCTCAACCCGATACAGAGCTGTGCGG 180
DB 62489 ATCTTCTACTGAGTGAAGCGCTCTGTGCTCTCAACCCGATACAGAGCTGTGCGG 62548
OY 181 GTGTTCCGCGAAGCGCTTTCGCGGTCACTGAGAAAGTGGGAATCAGCGCGGAGTAC 240
|||||

DB 62549 GTGTTCCGCGAAGCGCTTTCGCGGTCACTGAGAAAGTGGGAATCAGCGCGGAGTAC 62608
OY 241 TCGTCGGCATTTCCCGAGCTCAGCGATATGAAGATACGATTTGTCGGGCTGCCGCG 300
DB 62609 TCGTCGGCATTTCCCGAGCTCAGCGATATGAAGATACGATTTGTCGGGCTGCCGCG 62668
OY 301 GAGATCAACGCTCGAGTTCGAGAGCTGTCAACCCGTCGTTTACGTACCGCGCATTCGAC 360
DB 62669 GAGATCAACGCTCGAGTTCGAGAGCTGTCAACCCGTCGTTTACGTACCGCGCATTCGAC 62728
OY 361 CTGTCGCGCGCGGAAATACAGCGGACCGGTGACACAGCTGCTGATGCTGCTCCGGAACA 420
DB 62729 CTGTCGCGCGCGGAAATACAGCGGACCGGTGACACAGCTGCTGATGCTGCTCCGGAACA 62788
OY 421 GAGAGTTGCACTGTGTGCGGATTTACGCGGAGGAAATCCCGATGCGCGCATACGCGCT 480
DB 62789 GAGAGTTGCACTGTGTGCGGATTTACGCGGAGGAAATCCCGATGCGCGCATACGCGCT 62848
OY 481 CTGTTGAAGGTTCCGGCGAGTGTGACGAGAAAGTTCCGTCGCTTCGCTCGGCACTGCG 540
DB 62849 CTGTTGAAGGTTCCGGCGAGTGTGACGAGAAAGTTCCGTCGCTTCGCTCGGCACTGCG 62908
OY 541 CGCGCGCTCGCGGCTGTTGGTGGTGGTCCCGAGGTGATGAGAGACCAAGACCTGTCGCG 600
DB 62909 CGCGCGCTCGCGGCTGTTGGTGGTGGTCCCGAGGTGATGAGAGACCAAGACCTGTCGCG 62968
OY 601 TCCGTCACCGAGGGGCTCGCGCTGCTCATGACGTCCTCATGAGCGGCGAGGAACCCG 660
DB 62969 TCCGTCACCGAGGGGCTCGCGCTGCTCATGACGTCCTCATGAGCGGCGAGGAACCCG 63028
OY 661 CTCGAAATGACGTCCTTGACGATGCTCTCAGGCGGAGCGGACGAGGAGCTGAGC 720
DB 63029 CTCGAAATGACGTCCTTGACGATGCTCTCAGGCGGAGCGGACGAGGAGCTGAGC 63088
OY 721 ACGAAGAGCTGTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 63089 ACGAAGAGCTGTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63148
OY 781 TACCTTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 63149 TACCTTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63208
OY 841 AAGCCGAGCCCTGCTCATAGAGAACGCGCTCATGAGAGTCTCCGCTTTCAGCAATATC 900
DB 63209 AAGCCGAGCCCTGCTCATAGAGAACGCGCTCATGAGAGTCTCCGCTTTCAGCAATATC 63268
OY 901 CTCGAATAGGAACCTGTCCGCTTCCCGAGGACGAGCTGTAGTACTGCGGGGACATCCATC 960
DB 63269 CTCGAATAGGAACCTGTCCGCTTCCCGAGGACGAGCTGTAGTACTGCGGGGACATCCATC 63328
OY 961 AAGAAAGGGAGATGCTCTTCTCTCTGATCCCGAGCGCCCTGTAGAGATGGAGCTGATTC 1020
DB 63329 AAGAAAGGGAGATGCTCTTCTCTCTGATCCCGAGCGCCCTGTAGAGATGGAGCTGATTC 63388
OY 1021 TCCAGGCCAGACGTGTGTGATGTGACGGGACACGGGCGCGAGCTCGCGTACGGTATGA 1080
DB 63389 TCCAGGCCAGACGTGTGTGATGTGACGGGACACGGGCGCGAGCTCGCGTACGGTATGA 63448
OY 1081 GGGCCCCATGTCTGCCCCGGGGGTGTCCCTGTGCTCGAGGCGGAGATCGCCGTGGGC 1140
DB 63449 GGGCCCCATGTCTGCCCCGGGGGTGTCCCTGTGCTCGAGGCGGAGATCGCCGTGGGC 63508
OY 1141 ACCATCTTCGGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATATACAC 1200
DB 63509 ACCATCTTCGGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATATACAC 63568
OY 1201 CCCGCGTTCGGGAACATGAAATCACTCAACGTCACTTGAAGCCCTCCAAAGCTGGATAG 1260
DB 63569 CCCGCGTTCGGGAACATGAAATCACTCAACGTCACTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 4
US-09-568-480-1

```
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-480-1

Query Match      100.0%; Score 1260; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9.9e-290;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACACAGAGAGCAGATCAGATGAGACGAGCCTGCTTTCAGTTCAGCCGTTTC
DB 62369 ATGACACAGAGAGCAGATCAGATGAGACGAGCCTGCTTTCAGTTCAGCCGTTTC
QY 61 GGGCGCTGGGTACCGCGAGAGACCCGTTCCCGCGATGAGCGCTTGAGAGAGCAACCCG
DB 62429 GGGCGCTGGGTACCGCGAGAGACCCGTTCCCGCGATGAGCGCTTGAGAGAGCAACCCG
QY 121 ATCTTCTACTGGGATGAGAGCGCTCTGAGGCTCTCAGCCGATACACGAGCTGAGGCG
DB 62489 ATCTTCTACTGGGATGAGAGCGCTCTGAGGCTCTCAGCCGATACACGAGCTGAGGCG
QY 181 GTGTTCCGCGAGCAGACGCTTCCGCGATCAGTCAAGAGAGTGGGAAATCGAGCGGAGTAC
DB 62549 GTGTTCCGCGAGCAGACGCTTCCGCGATCAGTCAAGAGAGTGGGAAATCGAGCGGAGTAC
QY 241 TGTGCGGCAATTCGCGAGCTCAGCGATATGAGAGATGAGATGTTGCGGCGTCCGCGCG
DB 62609 TGTGCGGCAATTCGCGAGCTCAGCGATATGAGAGATGAGATGTTGCGGCGTCCGCGCG
QY 301 GAGGATCAGCGCTCGGAGTCCGCAAGCTGTCACACCCGTCGTTACGTCACGCGCATCGAC
DB 62669 GAGGATCAGCGCTCGGAGTCCGCAAGCTGTCACACCCGTCGTTACGTCACGCGCATCGAC
QY 361 CTGCTGCGGCGGAAATATACAGCGCACCGTCGACCACTGCTCGATCTCGTCCGAGCA
DB 62729 CTGCTGCGGCGGAAATATACAGCGCACCGTCGACCACTGCTCGATCTCGTCCGAGCA
QY 421 GAGGATGAGAGCTTGTGCGGAGATTCAGCGGAGAGATCCGATGCGCGCATCAGCGCT
DB 62789 GAGGATGAGAGCTTGTGCGGAGATTCAGCGGAGAGATCCGATGCGCGCATCAGCGCT
QY 481 CTGTTAAGAGTTCCGCGCGAGTGTGACGAGAGATTCGCTCGCTCGGCTCGGCGATCGG
DB 62849 CTGTTAAGAGTTCCGCGCGAGTGTGACGAGAGATTCGCTCGCTCGGCTCGGCGATCGG
QY 541 CGGCGCTCGCGCGTGGGTTGTGCGCCAGGTGATGAGAGACCAAGACCTTGATCGCG
DB 62909 CGGCGCTCGCGCGTGGGTTGTGCGCCAGGTGATGAGAGACCAAGACCTTGATCGCG
QY 601 TCCGTACACGAGGAGGCTCGCGCTGCTCATGAGCTCTCGATGAGGCGGCGAGAACCCG
DB 62969 TCCGTACACGAGGAGGCTCGCGCTGCTCATGAGCTCTCGATGAGGCGGCGAGAACCCG
QY 661 CTGAAATATGACGTTTGAAGATGCTGCTTCAAGCGCGGCGAGCGGACGAGCTGAGC
DB 63028 CTGAAATATGACGTTTGAAGATGCTGCTTCAAGCGCGGCGAGCGGACGAGCTGAGC
```

```
DB 63029 CTGAAATATGACGTTTGAAGATGCTGCTTCAAGCGCGGCGAGCGGACGAGCTGAGC
QY 721 ACGAAGAGCTGCTGCGCTGCTGAGTATTCGCTGCTGAGCAACGATACCAAGATC
DB 63089 ACGAAGAGCTGCTGCGCTGCTGAGTATTCGCTGCTGAGCAACGATACCAAGATC
QY 781 TACCTTATCGGCTTGTGCTGCTCAACCTGCTGCGGCTGCGGAGCGCTGAGCTGATG
DB 63149 TACCTTATCGGCTTGTGCTGCTCAACCTGCTGCGGCTGCGGAGCGCTGAGCTGATG
QY 841 AAGCGGAGCGGCTCATGAGAGAACCGCTGATGAGTGTCTGCTGACATATC
DB 63209 AAGCGGAGCGGCTCATGAGAGAACCGCTGATGAGTGTCTGCTGACATATC
QY 901 CTCAGATATGAACTGTGCTGCTTCCGAGAGAGAGCTGAGTACTGCGGCGATCATC
DB 63269 CTCAGATATGAACTGTGCTGCTTCCGAGAGAGAGCTGAGTACTGCGGCGATCATC
QY 961 AAGAAAGGAGAGATGCTTCTCTCTGATCCCGAGGCGCTGAGAGATGGGACGTATTC
DB 63329 AAGAAAGGAGAGATGCTTCTCTCTGATCCCGAGGCGCTGAGAGATGGGACGTATTC
QY 1021 TCCAGCGCAGACGTTTGTATGTGCGAGCGGACACGCGGAGCGCTGCGTACGTTAGA
DB 63389 TCCAGCGCAGACGTTTGTATGTGCGAGCGGACACGCGGAGCGCTGCGTACGTTAGA
QY 1081 GCGCCCATGTCTGCTGCGGAGTGTCTTGTGCTGCTGAGGCGGAGATGCGCTGAGC
DB 63449 GCGCCCATGTCTGCTGCGGAGTGTCTTGTGCTGCTGAGGCGGAGATGCGCTGAGC
QY 1141 ACCATCTCCGAGGTTCCCGAGATGAGAGTGAAGAACTCCGCTGTTGATATACAC
DB 63509 ACCATCTCCGAGGTTCCCGAGATGAGAGTGAAGAACTCCGCTGTTGATATACAC
QY 1201 CCGCGCTCCGAGATGAAATCACTCAACGTCTTGAAGCCCTCCAAAGCTGATAG
DB 63569 CCGCGCTCCGAGATGAAATCACTCAACGTCTTGAAGCCCTCCAAAGCTGATAG

RESULT 5
; US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; CURRENT FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
; US-09-568-486-1

Query Match      100.0%; Score 1260; DB 4; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9.9e-290;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACACAGAGAGCAGATCAGATGAGACGAGCCTGCTTTCAGTTCAGCCGTTTC
DB 62369 ATGACACAGAGAGCAGATCAGATGAGACGAGCCTGCTTTCAGTTCAGCCGTTTC
QY 61 GGGCGCTGGGTACCGCGAGAGACCCGTTCCCGCGATGAGCGCTTGAGAGAGCAACCCG
DB 62429 GGGCGCTGGGTACCGCGAGAGACCCGTTCCCGCGATGAGCGCTTGAGAGAGCAACCCG
```



```
Db 62429 GGGCTGGGTACGCGAGAGACCGGTTCCCGGATCGAGCGCTGAGAGGCAACCCCC 62488
Qy 121 ATCTTCTAGGAGATGAAGGCGGCTCTGGGTCCTCAACCGATACACAGAGCTGTGGG 180
Db 62489 ATCTTCTAGGAGATGAAGGCGGCTCTGGGTCCTCAACCGATACACAGAGCTGTGGG 62548
Qy 181 GTGTTCCGCGAGCAACGCTTCGCGGTCACTCGAGAAGAGTGGAAATCGAGCGGAGTAC 240
Db 62549 GTGTTCCGCGAGCAACGCTTCGCGGTCACTCGAGAAGAGTGGAAATCGAGCGGAGTAC 62608
Qy 241 TCGTGGGCAATTCGCGAGCTCAAGCATATGAAGAAAGTATGTTGGTGGGCTGCCGCG 300
Db 62609 TCGTGGGCAATTCGCGAGCTCAAGCATATGAAGAAAGTATGTTGGTGGGCTGCCGCG 62668
Qy 301 GAGGATCAAGCTCGGGTCCGCAAGGCTCGCAACCGGTGTTACGCAAGCGCATCGAC 360
Db 62669 GAGGATCAAGCTCGGGTCCGCAAGGCTCGCAACCGGTGTTACGCAAGCGCATCGAC 62728
Qy 361 CTGCTCGCGCGCGAATAACAGCGCAACCGTCAACAGCTGCTCGATGCTCGCGAGCA 420
Db 62729 CTGCTCGCGCGCGAATAACAGCGCAACCGTCAACAGCTGCTCGATGCTCGCGAGCA 62788
Qy 421 GAGGATTCGAGCTGTGCGGAGTACGCGAGGAGGAAATCCGATGCGCGCATCAAGCT 480
Db 62789 GAGGATTCGAGCTGTGCGGAGTACGCGAGGAGGAAATCCGATGCGCGCATCAAGCT 62848
Qy 481 CTGTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGCTGGGCTCGCGGACTGCG 540
Db 62849 CTGTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGCTGGGCTCGCGGACTGCG 62908
Qy 541 CGCGCGCTCGCGGTGGTGGTGGCCCGAGTCAATGAGAGAACCAAGCCCTGGTCCG 600
Db 62909 CGCGCGCTCGCGGTGGTGGTGGCCCGAGTCAATGAGAGAACCAAGCCCTGGTCCG 62968
Qy 601 TCCGTCACGAGAGGCGTCCGCTGCTCAATGAGCTCTGATGAGGCGCGCAGAACCG 660
Db 62969 TCCGTCACGAGAGGCGTCCGCTGCTCAATGAGCTCTGATGAGGCGCGCAGAACCG 63028
Qy 661 CTCGAAATGAGCTCTGAGAGTGTGCTCAAGCGCGAGCGAGCGGCGAGCTGAGC 720
Db 63029 CTCGAAATGAGCTCTGAGAGTGTGCTCAAGCGCGAGCGAGCGGCGAGCTGAGC 63088
Qy 721 ACGAAGAGCTGTGCGCGCTCGTGGGTGATTAATGCTGTGAGCAACGATACAGATC 780
Db 63089 ACGAAGAGCTGTGCGCGCTCGTGGGTGATTAATGCTGTGAGCAACGATACAGATC 63148
Qy 781 TACCTTATCGCTGTGCTGCTCAACCTGCTGCGGTGCGCGAGCGCTCGAGCTGTG 840
Db 63149 TACCTTATCGCTGTGCTGCTCAACCTGCTGCGGTGCGCGAGCGCTCGAGCTGTG 63208
Qy 841 AAGGCGAGCGCGGCGCTCAATGAGAAACGCGCTCGATGAGAGTCTCCGTTCAACAATC 900
Db 63209 AAGGCGAGCGCGGCGCTCAATGAGAAACGCGCTCGATGAGAGTCTCCGTTCAACAATC 63268
Qy 901 CTCAGAATAGAACTGTGCGTTCGCGAGGCAAGCACTGAGATGCTGCGGCGCATCGATC 960
Db 63269 CTCAGAATAGAACTGTGCGTTCGCGAGGCAAGCACTGAGATGCTGCGGCGCATCGATC 63328
Qy 961 AAGAAAGGAGATGCTTCTTCTCTGATCCGAGCGCTCGAGAGATGGAAGTGTATTC 1020
Db 63329 AAGAAAGGAGATGCTTCTTCTCTGATCCGAGCGCTCGAGAGATGGAAGTGTATTC 63388
Qy 1021 TCCAGGCGAGAGTGTGATGATGAGAGGAGCAAGGCGGAGAGCTCGCGATCAGTGA 1080
Db 63389 TCCAGGCGAGAGTGTGATGATGAGAGGAGCAAGGCGGAGAGCTCGCGATCAGTGA 63448
Qy 1081 GGGCCCCCATGTCTGCGCGGAGTGTCTCTGCTCGCTCGAGGCGAGATCGCGGTGAC 1140
Db 63449 GGGCCCCCATGTCTGCGCGGAGTGTCTCTGCTCGCTCGAGGCGAGATCGCGGTGAC 63508
Qy 1141 ACCATCTTCGTAAGTTCCCGAGATGAAGCTGAAGAAACTCCGCTGTTGGATACAC 1200
```

```
Db 63509 ACCATCTTCGTAAGTTCCCGAGATGAAGCTGAAGAAACTCCGCTGTTGGATACAC 63568
Qy 1201 CCCGCGTCCGGAACATGATCACTCAAGCTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
Db 63569 CCCGCGTCCGGAACATGATCACTCAAGCTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-10582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-472-1
```

Query Match 100.0%; Score 1260; DB 4; Length 68750;

Best Local Similarity 100.0%; Pred. No. 9.9e-290; Mismatches 0; Indels 0; Gaps 0;

Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ATGACACAGAGCAACGAAATCAGATGAGACGAAAGCTGCTTGTGACTTCAAGCGTTTC 60
Db 62369 ATGACACAGAGCAACGAAATCAGATGAGACGAAAGCTGCTTGTGACTTCAAGCGTTTC 62428
Qy 61 GCGCTGGGTACGCGGAGAGACCCGTTCCCGGATCGAGCGCTGAGAGAGGCAACCC 120
Db 62429 GCGCTGGGTACGCGGAGAGACCCGTTCCCGGATCGAGCGCTGAGAGAGGCAACCC 62488
Qy 121 ATCTTCTAGGAGATGAAGGCGGCTCTGGGTCTCAACCGATACACAGAGTGTGGG 180
Db 62489 ATCTTCTAGGAGATGAAGGCGGCTCTGGGTCTCAACCGATACACAGAGTGTGGG 62548
Qy 181 GTGTTCCGCGAGCAACGCTTCGCGGTCACTCGAGAAGAGTGGAAATCGAGCGGAGTAC 240
Db 62549 GTGTTCCGCGAGCAACGCTTCGCGGTCACTCGAGAAGAGTGGAAATCGAGCGGAGTAC 62608
Qy 241 TCGTGGGCAATTCGCGAGCTCAAGCATATGAAGAAAGTATGTTGGTGGGCTGCCGCG 300
Db 62609 TCGTGGGCAATTCGCGAGCTCAAGCATATGAAGAAAGTATGTTGGTGGGCTGCCGCG 62668
Qy 301 GAGGATCAAGCTCGGGTCCGCAAGGCTCGCAACCGGTGTTACGCAAGCGCATCGAC 360
Db 62669 GAGGATCAAGCTCGGGTCCGCAAGGCTCGCAACCGGTGTTACGCAAGCGCATCGAC 62728
Qy 361 CTGCTCGCGCGCGAATAACAGCGCAACCGTCAACAGCTGCTCGATGCTCGCTCGGACAA 420
Db 62729 CTGCTCGCGCGCGAATAACAGCGCAACCGTCAACAGCTGCTCGATGCTCGCTCGGACAA 62788
Qy 421 GAGGATTCGAGCTGTGCGGAGTACGCGAGGAGGAAATCCGATGCGCGCATCAAGCT 480
Db 62789 GAGGATTCGAGCTGTGCGGAGTACGCGAGGAGGAAATCCGATGCGCGCATCAAGCT 62848
Qy 481 CTGTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGCTGGGCTCGCGGACTGCG 540
Db 62849 CTGTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGCTGGGCTCGCGGACTGCG 62908
Qy 541 CGCGCGCTCGCGGTGGTGGTGGCCCGAGTCAATGAGAGAACCAAGCCCTGGTCCG 600
```

Db	62909	CGCGCGCTCGGCGGTGGGTTTGGTGGCCCAAGTGGATGAGGAGACCAAGACCCCTGGTGGG	62968
Qy	601	TCGCTCACCCGAGGGGGCTCGCGCTGCTCATGACGTCTCTGATGAGCGGGCAGAGAACCCG	660
Db	62369	TCCCTCACCCAGGGGGGCTCGCGCTGCTCATGAGTCTCTCATGATGAGGCGCAGAGAACCCG	63028
Qy	661	CTGCAAAATGACGCTTTGACGATGCTGCTTCAGGCGGAGGCGCAGCGACAGCGCTGAGC	720
Db	63029	CTCGAAAATGACGCTTTGACGATGCTGCTTCAGGCGGAGGCGCAGCGACAGCGCTGAGC	63088
Qy	721	ACGAGGAGCGTGTGCTCGCGCTCGTGAGGTGCGATTATGCTGCTTGACAACGATACGATC	780
Db	63089	ACGAGGAGCGTGTGCTCGCGCTCGTGAGGTGCGATTATGCTGCTTGACAACGATACGATC	63148
Qy	781	TACCTTATCGCGTTTCCGTGCTGCTTCACCTGCTGCGGTGCGCCGAGGCGCTCCAGCTGGTG	840
Db	63149	TACCTTATCGCGTTTCCGTGCTGCTTCACCTGCTGCGGTGCGCCGAGGCGCTCCAGCTGGTG	63208
Qy	841	AAGGCCGAGCCCGGGGCTCATGAGGAGACGCGCTGATGAGGTGCTCCGCTTCGACAAATATC	900
Db	63209	AAGGCCGAGCCCGGGGCTCATGAGGAGACGCGCTGATGAGGTGCTCCGCTTCGACAAATATC	63268
Qy	901	CTCAGAAATAGAACTGTGCTGCTTTCGCCAGGACAGACCTGAGATGACTCGGGGATGATC	960
Db	63269	CTCAGAAATAGAACTGTGCTGCTTTCGCCAGGACAGACCTGAGATGACTCGGGGATGATC	63328
Qy	961	AAGAAAGGGGAGATGAGTCTTTCTCTGATATCCGAGGCGCCCTGAGAAATGGGACTGATATC	1020
Db	63329	AAGAAAGGGGAGATGAGTCTTTCTCTGATATCCGAGGCGCCCTGAGAAATGGGACTGATATC	63388
Qy	1021	TCCAGGCGCAGACGCTGTTGATGATGCGACGCGGACACGCGGCGCGAGCCTCGCTACGGTAGA	1080
Db	63389	TCCAGGCGCAGACGCTGTTGATGATGCGACGCGGACACGCGGCGCGAGCCTCGCTACGGTAGA	63448
Qy	1081	GGCCCCCATGTCGCCCCGGGGGTGTCCTTGTCTCGCTCGAGGCGGAGATCGCCGTGGGC	1140
Db	63449	GGCCCCCATGTCGCCCCGGGGGTGTCCTTGTCTCGCTCGAGGCGGAGATCGCCGTGGGC	63508
Qy	1141	ACCATCTTCCGTAGGTTCCCGGAGATGAACTGAAAGAAATCCCGCTGTTGGATTACAC	1200
Db	63509	ACCATCTTCCGTAGGTTCCCGGAGATGAACTGAAAGAAATCCCGCTGTTGGATTACAC	63568
Qy	1201	CCCCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCAAAAGCTGGATAG	1260
Db	63569	CCCCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCAAAAGCTGGATAG	63628
RESULT 7			
US-09-567-899-1			
: Sequence 1, Application US/09567899			
: Patent No. 6383787			
: GENERAL INFORMATION:			
: APPLICANT: Schnupp, Thomas			
: APPLICANT: Ligon, James			
: APPLICANT: Molnar, Istvan			
: APPLICANT: Zirkle, Ross			
: APPLICANT: Cyr, Devon			
: APPLICANT: Goerlach, Joern			
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES			
: FILE REFERENCE: 4-30582A			
: CURRENT APPLICATION NUMBER: US/09/567,899			
: CURRENT FILING DATE: 2000-05-10			
: PRIOR APPLICATION NUMBER: 09/335,409			
: PRIOR FILING DATE: 1999-06-17			
: NUMBER OF SEQ ID NOS: 30			
: SOFTWARE: PatentIn Ver. 2.0			
: SEQ ID NO 1			
: LENGTH: 68750			
: TYPE: DNA			
: ORGANISM: Sorangium cellulosum			
US-09-567-899-1			

Query Match		100.0%;	Score 1260;	DB 4;	Length 68750;
Best Local Similarity		100.0%;	Pred. No. 9.9e-290;		
Matches 1260;	Conservative	0;	Mismatches	0;	Indels
0;	Gaps	0;			
OY	1	ATGACACAGGAGCAGCAATCAAGATGAGACGAAGACCTCTGTTTGACTTCAGACCGCTTC	60		
Db	62369	ATATACACAGAGAGCAGACGAAATCAGATGAGAGGAAGACCTCTGTTTGACTTCAGACCGCTTC	62428		
OY	61	GCGCGCTGGGATCGCGAGAGACCGCTTCCCGCGATCGAGCGCTGAGAGGCAACCCG	120		
Db	62429	GCGCGCTGGGATCGCGAGAGACCGCTTCCCGCGATCGAGCGCTGAGAGGCAACCCG	62488		
OY	121	ATCTTCTACTGGAGATGAGCGCGCTCTGGGTCTTCACCCGATACACGACGTGTGCGG	180		
Db	62489	ATCTTCTACTGGAGATGAGCGCGCTCTGGGTCTTCACCCGATACACGACGTGTGCGG	62548		
OY	181	GTTGTCGCGAGACGAAAGCTTCCGCGGTGATCGAGAAAGTGGGATTCGAGCGCGAGATC	240		
Db	62549	GTTGTCGCGAGACGAAAGCTTCCGCGGTGATCGAGAAAGTGGGATTCGAGCGCGAGATC	62608		
OY	241	TCTGTGGCCATTCCCGAGCTCAGACGATATGAAAGAGTACGATTTGGGTGCGCGCG	300		
Db	62609	TCTGTGGCCATTCCCGAGCTCAGACGATATGAAAGAGTACGATTTGGGTGCGCGCG	62668		
OY	301	GAGGATCACGCTTGGGTCGCGAAGCTGTGCAACCGGTGTTTACGTACGGCGCATCGAC	360		
Db	62669	GAGGATCACGCTTGGGTCGCGAAGCTGTGCAACCGGTGTTTACGTACGGCGCATCGAC	62728		
OY	361	CTGCTGCGCGCCGAATATACGCGCACCGCTGACACAGCTGCTGAGTCTGCTCCGAGAA	420		
Db	62729	CTGCTGCGCGCCGAATATACGCGCACCGCTGACACAGCTGCTGAGTCTGCTCCGAGAA	62788		
OY	421	GAGAGATTGACGTTGTGCGGGATTACGCGAGGGAAATCCGATGCGCGCATACGCGCT	480		
Db	62789	GAGAGATTGACGTTGTGCGGGATTACGCGAGGGAAATCCGATGCGCGCATACGCGCT	62848		
OY	481	CTGTGAAGGTTCCGGCCGAGGTGACGAGAAAGTTCCGTGCGTTCGGGCTCGCGCATGCG	540		
Db	62849	CTGTGAAGGTTCCGGCCGAGGTGACGAGAAAGTTCCGTGCGTTCGGGCTCGCGCATGCG	62908		
OY	541	CGCGCGCTCGCGCTGGGTTTGTGTCGCCCAAGTCGATGAGAGACCAAGACCTGTGTCG	600		
Db	62909	CGCGCGCTCGCGCTGGGTTTGTGTCGCCCAAGTCGATGAGAGACCAAGACCTGTGTCG	62968		
OY	601	TCCGTACACGAGGGGCTTCGCGCTGTCTCATGACGTCCTGATGAGACCGCGCAGAAACCG	660		
Db	62969	TCCGTACACGAGGGGCTTCGCGCTGTCTCATGACGTCCTGATGAGACCGCGCAGAAACCG	63028		
OY	661	CTGGAATAATGACGCTTTGACGATGCTGCTTACGCGCAGCGCAGCGCAGCGACGCTGAGC	720		
Db	63029	CTGGAATAATGACGCTTTGACGATGCTGCTTACGCGCAGCGCAGCGCAGCGACGCTGAGC	63088		
OY	721	ACGAGAGAGCTGTGCTCGCGCTCGTGGGTGCGATTATCGCTGTGCGACCGGATCCACGATC	780		
Db	63089	ACGAGAGAGCTGTGCTCGCGCTCGTGGGTGCGATTATCGCTGTGCGACCGGATCCACGATC	63148		
OY	781	TACCTTATCGCGTTTCGCTGTGCTCAACGCTGCGCGGTGCGCCGAGCGCTCGAGCTGGTG	840		
Db	63149	TACCTTATCGCGTTTCGCTGTGCTCAACGCTGCGCGGTGCGCCGAGCGCTCGAGCTGGTG	63208		
OY	841	AAGGCGCAGGCCCGGGCTCATGAGGAAGCGGCTCGATGAGGTGCTCCGTTGCAAAATC	900		
Db	63209	AAGGCGCAGGCCCGGGCTCATGAGGAAGCGGCTCGATGAGGTGCTCCGTTGCAAAATC	63268		
OY	901	CTCAGAAATAGGAACGTGTGCTTTCGCGCAGGACGAACTTGGAATATCGCGGGGCAATCGATC	960		
Db	63269	CTCAGAAATAGGAACGTGTGCTTTCGCGCAGGACGAACTTGGAATATCGCGGGGCAATCGATC	63328		
OY	961	AAGAAAGGGAGATGTGCTTCTCTCGATATCCGAGCGCCCTGAGAGATGGACTGTATTC	1020		
Db	63329	AAGAAAGGGAGATGTGCTTCTCTCGATATCCGAGCGCCCTGAGAGATGGACTGTATTC	63388		
OY	1021	TTCAGGCGACAGCTGTTTGAATGTGCGACGGAGACCGGCGCGAGCTCCGCTACGATAG	1080		

Db 63389 TCACGCGCAGACGCTGTTGATGATGAGACGAGCAGCGGCGAGCTGCGTACGGTGA 63448
Qy 1081 GGGCCCCCATGCTGTCGCCCCGGGGGTGTCCTGTGCTGCGCTCGAGGGGAGATGCGCGTGGGC 1140
Db 63449 GGGCCCCCATGCTGTCGCCCCGGGGGTGTCCTGTGCTGCGCTCGAGGGGAGATGCGCGTGGGC 63508
Qy 1141 ACCATCTTCGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATACAC 1200
Db 63509 ACCATCTTCGTAGGTTCCCGAGATGAAGCTGAAGAACTCCCGTGTGGATACAC 63568
Qy 1201 CCGCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1260
Db 63569 CCGCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 63628

RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kusan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT FILING DATE: US/09/443, 501A
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 98.9%; Score 1246.2; DB 4; Length 71989;
Best Local Similarity 99.4%; Pred. No. 1.9e-286;
Matches 1251; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGACACAGAGCAGCAAGATAGAGTGAAGCAAGCTGCTTTCGATTCGAAGCCGTTTC 60
Db 56757 ATGACACAGAGCAGCAAGATAGAGTGAAGCAAGCTGCTTTCGATTCGAAGCCGTTTC 56816
Qy 61 GCGCGCTGGTACGCGGAGAGCCGTTCCCGCGATCGAGCGCTTGAAGAGGCAACCCCC 120
Db 56817 GCGCGCTGGTACGCGGAGAGCCGTTCCCGCGATCGAGCGCTTGAAGAGGCAACCCCC 56876
Qy 121 ATCTTCTACTGGAGTGAAGGCGCTCCTGGGTCTCAACCGATTCACAGAGTGTGCGCG 180
Db 56877 ATCTTCTACTGGAGTGAAGGCGCTCCTGGGTCTCAACCGATTCACAGAGTGTGCGCG 56936
Qy 181 GTGTTCCGCGAGCAGACGCTTCCGCTGCTGAGAGAGTGGAGATCGAGCGCGAGTAC 240
Db 56937 GTGTTCCGCGAGCAGACGCTTCCGCTGCTGAGAGAGTGGAGATCGAGCGCGAGTAC 56996
Qy 241 TCGTTCGCGCATTCGCGAGCTCAAGCATATGAAGAGTGAAGTGTGCGCGTGCCTGCGCG 300
Db 56997 TCGTTCGCGCATTCGCGAGCTCAAGCATATGAAGAGTGAAGTGTGCGCGTGCCTGCGCG 57056

Qy 301 GAGATACAGCTCGGCTCCGAGACCTGTCACACCCGTCGTTACGTCACGCGCATCGAC 360
Db 57057 GAGATACAGCTCGGCTCCGAGACCTGTCACACCCGTCGTTACGTCACGCGCATCGAC 57116
Qy 361 CTGCTGGCGCGCGAAATACAGCCGACCGTGCACCAAGCTGCTGATGCTGCTCCGAGACA 420
Db 57117 CTGCTGGCGCGCGAAATACAGCCGACCGTGCACCAAGCTGCTGATGCTGCTCCGAGACA 57176
Qy 421 GAGAGTTCGACGTTGTGCGGAGTTACCGGAGGAAATCCGAGTCCGCGCATCGAGCT 480
Db 57177 GAGAGTTCGACGTTGTGCGGAGTTACCGGAGGAAATCCGAGTCCGCGCATCGAGCT 57236
Qy 481 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCGCTTCGCTTCGCGTCCGACATCGC 540
Db 57237 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCGCTTCGCTTCGCGTCCGACATCGC 57296
Qy 541 CCGCGCTCGCGCTGGTGTGTCGCCAGGTCCAGGTGCATGAGAGACCAAGACCTGGTCCG 600
Db 57297 CCGCGCTCGCGCTGGTGTGTCGCCAGGTCCAGGTGCATGAGAGACCAAGACCTGGTCCG 57356
Qy 601 TCGCTACCGAGGGGCTCGCGCTGCTCATGACGTCCTCGATGAGCGCGCGAGAACCG 660
Db 57357 TCGCTACCGAGGGGCTCGCGCTGCTCATGACGTCCTCGATGAGCGCGCGAGAACCG 57416
Qy 661 CTCGAAATGACGCTCTTGACGATGCTGCTTCAGGCCGAGCGCGACGAGCGCTGAGC 720
Db 57417 CTCGAAATGACGCTCTTGACGATGCTGCTTCAGGCCGAGCGCGACGAGCGCTGAGC 57476
Qy 721 ACGAAGAGCTGTGCGCGCTGCGGTCGATTAATGCTGTCGACCCAGATACAGATC 780
Db 57477 ACGAAGAGCTGTGCGCGCTGCGGTCGATTAATGCTGTCGACCCAGATACAGATC 57536
Qy 781 TACCTTATCGGTTCCGCTGCTGCTCAACCTGTGCGGTGCGCGAGCGCTGAGCTGCTG 840
Db 57537 TACCTTATCGGTTCCGCTGCTGCTCAACCTGTGCGGTGCGCGAGCGCTGAGCTGCTG 57596
Qy 841 AAGCCGAGCCCGGCTCATGAGAGACGCGCTCATGAGAGTGTCTCCGCTTCGACAAATATC 900
Db 57597 AAGCCGAGCCCGGCTCATGAGAGACGCGCTCATGAGAGTGTCTCCGCTTCGACAAATATC 57656
Qy 901 CTCGAAATGAGAACTGTGCGCTTTCGCGAGCGAGACCTGAGATCTGCGGCGCATTCGATC 960
Db 57657 CTCGAAATGAGAACTGTGCGCTTTCGCGAGCGAGACCTGAGATCTGCGGCGCATTCGATC 57716
Qy 961 AAGAAAGGAGAGTGTCTTCTCTGATCCCGAGCGCCCTGAGAGATGGAGCTGATTC 1020
Db 57717 AAGAAAGGAGAGTGTCTTCTCTGATCCCGAGCGCCCTGAGAGATGGAGCTGATTC 57776
Qy 1021 TCCAGGCCAGACGTTTGAATGTGCGACGAGGACAGCGGCGCGAGCTTCGCTACGATGAGA 1080
Db 57777 TCCAGGCCAGACGTTTGAATGTGCGACGAGGACAGCGGCGCGAGCTTCGCTACGATGAGA 57836
Qy 1081 GGGCCCATGCTGTCGCGGGGTGTCCTTGTGCTGCGCTCGAGGCGAGATGCGGTGGGC 1140
Db 57837 GGGCCCATGCTGTCGCGGGGTGTCCTTGTGCTGCGCTCGAGGCGAGATGCGGTGGGC 57896
Qy 1141 ACCATCTTCGTAGGTTCCCGAGATGAAGTGAAGAAATCTCCGCTGTGGATACAC 1200
Db 57897 ACCATCTTCGTAGGTTCCCGAGATGAAGTGAAGAAATCTCCGCTGTGGATACAC 57956
Qy 1201 CCGCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 1259
Db 57957 CCGCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATAG 58015

RESULT 9
US-09-252-991A-1265
; Sequence 1265, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 1265
LENGTH: 1314
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1265

Query Match 11.5%; Score 145.4; DB 4; Length 1314;
Best Local Similarity 48.8%; Pred. No. 1.1e-25;
Matches 456; Conservative 0; Mismatches 446; Indels 33; Gaps 1;

254 CCGAGCTCAGCATATGAAAGTAAGATTGTCGGGCTGCCCGCGAGATCAGCTC 313
DB CCGGATTCGGCGAGGAGCTCAGCCACACATGCTCAACTCGACCGCGGACATATCC 349
314 GGGTCGCAAGCTCGTCAACCCGTCGTTACGTCAGCGCCATCGACTGTCGGCGCG 373
DB GCGTCGCTCGCTGCTGCTGCGCGCTTCAACCCGCGGCGAGTGGAGCCGCTGCAACCCG 409
374 AATAACAGCGACCGTGCACAGCTGCTGATGCTGCTCCCGACAAGAGATTGACG 433
DB ATATGAAAGGATCAACCGAGAAATTGCTGAGCCATGAGCCGCGGCAACAGCGGACC 469
434 TTGTGGGATTAACGGAGGGAATCCGATGGCGCGATCAAGCGCTGTTGAAGTTC 493
DB TATAGCCGACTTCGATGCCGCTGACATGCGGTGATCTTGAGCTGCTGGGCAATTC 529
494 CGGCGAGTGTACAGAAAGTTCCGTCGCTCGGCTCGGCGACTGCGCGCGCTCGCGG 553
DB CGAGGCGGAGCGGCAACGCGCGGCACTCTG----- 563
554 TGGGTTTGGTCCCGAGTGCATGAGAGACCAAGACCTGTGCGCTGCTGACCGAGG 613
DB -----GGAGGCGCGAGCGGAACTGCTGCGCGAGAGGCGCGCGCTGCGCGAGT 616
614 GCGTCGCGCTGCTCATGAGCTGCTGATGAGCGGCGAGAAACCGCTCGAAAAAGAG 673
DB CCGAGTGCATCACTGCGCGGTGCTGCTGAGGCGCAAGCGCGCGGCGAGCGGAGG 676
674 TCTTGCATGATGCTTCAAGCGGAGCGCGAGCGGCTGAGACAGAAAGAGCTGG 733
DB TCTACAGCGGCTGCTGAGCGCGCGGAGAGCGGCGCAATGAGCGAGCGGAGCTGG 736
734 TCGCGCTCGTGGGTGATTAATGCTGCTGAGCGGAGTCAACGATCTTAATGCGCT 793
DB TCTCCATGCGCAACCTGCTGATGATGAGCGGCTTCAAGACCAATGAACATGATGGGA 796
794 TGGGTTGCTCAACCTGCTGCGGTGCGCGGCGCTGAGCTGAGGAGCGGAGCGG 853
DB AGCGCTGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856
854 GGGTCATGAGAAAGCGGCTGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 913
DB AACTCTGCGCAACCGCATGAAAGAACTGCTGCGGCAAGCGCGGCTGCGGCTGCA 916
914 CTGTGCGTTTCCGCGAGGAGCTGAGATCTGCGGCGGCTGATCAAGAAAGGAGGA 973
DB TGTGCGCTTCACTGCGGAGAGCTGAGAACTGAGCGGCTGCTGCTGCGCGGAGAT 976
974 TGGTCTTCTGCTGATCCGAGCGGCGCTGAGAGTGGAGCTGATCTTCAAGCGGAGAG 1033
DB AATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1036
1034 TGTGATGTGCGAGCGGAGCAGCGGCGGAGCTGCGCTGATGAGGCGCGGCTGCTGCT 1093

DB 1037 GCGTCAGCTCAGCCGCAACCGATGGCATCTGCGTACGCGCTTGGCGGCACTACT 1096
QY 1094 GCGCGGAGTGTCTTCTGCTGCTGCGGAGGAGATGCCGTGGGCACTTCTCGTA 1153
DB 1097 GCGTCGCGGCTGCTGCGGCGGAGGAGGCGGATGCCATTCAGCGGCTGCTGCGG 1156
QY 1154 GGTTCGCGAGTGAAGCTGAAGAACTCCCGT 1188
DB 1157 GCTTCCGACTCCAGTTGGCGGTGCCCAAGG 1191

RESULT 10
US-09-252-991A-1338/c
Sequence 1338 Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 1338
LENGTH: 1482
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1338

Query Match 11.5%; Score 145.4; DB 4; Length 1482;
Best Local Similarity 48.8%; Pred. No. 1.2e-25;
Matches 456; Conservative 0; Mismatches 446; Indels 33; Gaps 1;

254 CCGAGCTCAGCATATGAAAGTAAGATTGTCGGGCTGCCCGGAGATCAGCTC 313
DB 1217 CCGGATTCGGCGAGGAGCTCAGCCACACATGCTCAACTCGACCGCGGACATATCC 1158
314 GGGTCGCAAGCTCGTCAACCCGTCGTTACGTCAGCGGCGCATCGACTGTCGCGCG 373
DB 1157 GCGTCGCTCGTGTGCTGCTGCGCGGCTTCAACCCGCGGCGAGTGGAGCGGCTGCAACCG 1098
374 AATAACAGCGACCGTGCACAGCTGCTGATGCTGCTCGGAGACAAGAGATTGACG 433
DB ATATGAAAGGATCAACCGAGAAATTGCTGAGCGCCATGCGCGCGCGCAACAGCGGACC 1038
434 TTGTGCGGATTAACGGAGGGAATCCGATGGCGGCGGATGAGCGCTGTTGAAGTTC 493
DB TATAGCCGACTTCGATGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978
494 CCGCGAGTGTACAGAAAGTTCCGTCGCTTCCGCTCGGCGACTGCGCGCGCTCGGCG 553
DB CCGAGGCGGAGCGGAAACGCGCGGCACTCTG----- 944
554 TGGGTTGCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 613
DB 943 -----GGAGGCGCGAGCGGAACTGCTGCTGCGCGGAGAGCGGCGGCTGCGGAG 891
614 GCGTCGCGCTGCTCATGAGCTGCTGATGAGGAGCGGAGAAACCGCTGCAAAATGAG 673
DB CCGAGTGCATCACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 831
674 TCTTGCATGATGCTTCAAGCGGAGCGGAGCGGAGGCTGAGCAAGAGAGCTGG 733
DB TCTACAGCGGCTGCTGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 771
734 TCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 793
DB TCTCCATGCGCAACCTGCTGATGATGAGCGGCTTGAAGCAACATGAACATGATCGGA 711

```

Oy 794 TCGCTGCTCAACCTGCTGCGGTGCGCCGAGGCGCTCGAGCTGTGAGAGCCGAGCCG 853
Db 710 ACGGCTGTGTCAACCTGCTGCTGCTCAACCCGAGACCACTGAGCTGTGCGGCGAGCCG 651
Oy 854 GGCCTCATAGAGAACCGCGCTCATGAGTGTCTCCGCTTGACCAATCTCTCAGAAATAGAA 913
Db 650 AACTCTGCGCCCAACCGCATGAGAACTGTGCTCCGCAACAGCCCGGTGCGCGCTTGA 591
Oy 914 CTGTGCGTTTGCGCCGAGGAGGAGCTGAGTACTGTGCGGGGCATTCGATCAAGAAAGGAGAG 973
Db 590 TGTTCGCTTCAACCTGAGAAAGCTGGAATGGAAGGAGGTCACTTCTCCGCGGCAAT 531
Oy 974 TGTCTTCTCTCTGATCCGAGCGCCCTGAGAGATGGAATGTATCTTCAGAGCCAGAG 1033
Db 530 ACATCTGCTGTCTCCAACTGAGCCGCAACAGAGCGCGAGCGCTTGACAGATCCGAGCC 471
Oy 1034 TGTTCGATGTGAGAGGAGACAGCGGCGAGCGCTGCGTACGCTGAGAGCCCGCATGTCT 1093
Db 470 GCTTCGACCTCAACCGCAACACCGATGCGCATCTCGGCTACGCGCTTCGCGCTGCACTACT 411
Oy 1094 GCGCGGCGGTGTCTGCTGCTGCGCTCGAGGCGAGATGCGCGTGGGCAATCTTCCGTA 1153
Db 410 GCGTGGCGGCTCTGCTGCGCGCGCTGAGGCGGAGTGCATCCAGCGCTGCTGCGCC 351
Oy 1154 GGTTCGCCGAGATGAAAGCTGAAAGAACTCCCGTG 1188
Db 350 GCTTCCCGGACCTCCAGTTGGCGGTGCCCAAGCG 316

```

```

RESULT 11
US-09-252-991A-1222
; Sequence 1222, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rudenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1222
; LENGTH: 1671
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1222

```

```

Query Match 11.5%; Score 145.4; DB 4; Length 1671;
Best Local Similarity 48.8%; Pred. No. 1.2e-25;
Matches 456; Conservative 0; Mismatches 446; Indels 33; Gaps 1;

```

```

Oy 254 CCGAGCTCAGCGATATGAAGAATGATGATGTTGGGCTGCGCGGAGGATCAGCGTC 313
Db 225 CCGGATGAGGAGGAGATCTGACCCACCAATGCTCACTGACCCGCGGAGCACTACCC 284
Oy 314 GGGTCCGCAAGCTGCTCAACCCGCTTACGTCACGCGCATTCAGCTGTGCGCGCG 373
Db 285 GCTGCGCTGCTGCTGCGCGCGCGGCTTACCCGCGCAAGTGAAGCGCTGCAACGCG 344
Oy 374 AAATACAGCCGAGCTGACCAAGCTGCTGATGCTGCTGCGGACAAAGAGATTGACG 433
Db 345 ATATAGAACGAGATCCAGAGAAATTGTGAGCGCATGCGCGCGCGCAACAGCGGAGCC 404
Oy 434 TTGTGCGGATTTACGCGAGGAAATCCGATGCGCGCATGAGCGCTCTGTTGAAGTTTC 493
Db 405 TGATGCGGACCTTCCGATCCCGCTGACCATGCGGATCTTTCAGAGTGTGCGGCAATTC 464
Oy 494 CGGCGAGTGTGACGAGAAGTTCCGTGCTTTCGCTGCGGCACTGCGCGCGCTGCGCG 553

```

```

Db 465 CCGAGGCGGAGCGGAACACGCGCCGCAATCTCTG----- 498
Oy 554 TGGGTTTGATGCCCAAGGTGATGAGAGAACCAAGACCTGTGCGCTGCTCACCGAG 613
Db 499 -----GAGCGCGAGCGGAATCTGCTGTGCGGAGAGAGGCCAGGCTTGTGCGCTAG 551
Oy 614 GGTCTGCGCTGCTCATGACGCTCTCGATGAGAGCGCGGAGAACCCGCTCGAAAAATGACG 673
Db 552 CGCAGGTGACTATCTGCGCGCTGTGCTGAGAGCGCAAGCGCGGAGCGACCGACGAGCG 611
Oy 674 TCTTGACGATGCTGCTTTCAGGCGCGAGCGGAGCGAGCTGAGACAGAAAGAGCTGCG 733
Db 612 TCTACAGCGGCTGTGTGAGGCGCGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCTGCG 671
Oy 734 TCGGCTGCTGAGGAGGATATTCGCTGCTGCGGACCGATACGACGATCACTTATCGGCT 793
Db 672 TCTCATGAGCCGACCTGTGTATGATGAGCGGCTTGTGAGACCACTGAACATGATCGGCA 731
Oy 794 TCGCTGCTCAACCTGCTGCGGTGCGCGCGAGGCGCTGAGCTGTGAGAGCCGAGCCG 853
Db 732 ACGGCTGTGTCAACCTGCTGCTGCAACCCGAGCAATGCGGCTTGTGCGGCGGAGCGCG 791
Oy 854 GGCCTCATAGAGAACGCGCTCATGAGAGTGTCTCCGCTTGACCAATATCTTCAGATAGAA 913
Db 792 AACTCTGCGCCCAACGCGATGAGAACTGTGCTCCGCAACAGCGCGGCTGCGCGCTGCA 851
Oy 914 CTGTGCGTTTGCGCAGGAGGACCTGAGAGTACTGTGCGGCGGATCGATCAAGAAAGGAGAG 973
Db 852 TGTTCGCTTACCTGTGAGAGAGCTGGAATGGAACGAGGAGTCACTTCCGCGGCGAAT 911
Oy 974 TGTCTTCTCTGATCCGAGCGCCCTGAGAGATGAGACTGTATTTCTCAGAGCCAGAG 1033
Db 912 ACATCTGCTGTCCAACTGACCGCAACAGCAACAGCGCGGAGGCTTGACATCCGAGCC 971
Oy 1034 TGTTCGATGTGAGAGGAGACAGGCGCGGAGCTGCTGCTGATGAGAGGCGCCCATGTCT 1093
Db 972 GCTTCGACCTCAACCGCAACCGATGCGCATCTTCGCTACGCGCTTCGCGCTGCACTACT 1031
Oy 1094 GCGCGGCGGTGTCTGCTGCTGCGCTGAGGCGGAGATGCGCGTGGGCACTCTTCCGTA 1153
Db 1032 GCGTGGCGCTCTGCTGCGCGCGGCTGAGGCGGAGATGCGCATCCAGCGCTGCTGCGCC 1091
Oy 1154 GGTTCGCCGAGATGAAAGCTGAAAGAACTCCCGTG 1188
Db 1092 GCTTCCCGGACCTCCAGTTGGCGGTGCCCAAGCG 1126

```

```

RESULT 12
US-09-105-537-38
; Sequence 38, Application US/09105537A
; Patent No. 6265302
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-38

```

```

Query Match 9.8%; Score 123; DB 3; Length 1251;
Best Local Similarity 47.5%; Pred. No. 2.3e-20;
Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;
Oy 236 AGTACTGCTGCGCATTCCTCCAGCTCAGCGATATGAGAAAGTACGATTTGTCGCGCTGC 295

```

```

Db      227 ACTCCAGCAGCTCCCTGACCGAGCGGAGCGCGCTCAACCAACATGCTGAGTCCG 286
Oy      296 CGCCGAGAGATCAAGCTCGGGTCCGCAAGCTGTCACACCGCTTACATGCAAGCGGCA 355
Db      287 ACCCGCGGCGGACACCGGCTGCGGACCTGGTGGCCGTGATGTTCAACATGCGCGG 346
Oy      356 TCGACCTGCTGGCGCGGAAATACAGCGACCTGTGACCAAGTGTCTGATGCTCGCTCG 415
Db      347 TCGAGTTGCTGGCGCGCGGCTGACAGAGATGTCACAGCGCTGTCGAGCGCATGCTG 406
Oy      416 GACAAGAGAGATTCAGCTTGTGCGGAGATTACGCGAGGAAATCCGATGCGCGCATCA 475
Db      407 CGGCGCGGCGGACCGCGCGCGCATCTGATGAGAGTCCCTGCGCTGCGCGCTGC---CGATCA 463
Oy      476 GCGCTCTGTGAAGTTCCGCGCGAGATGTCAGAGAAATTCGCTCGCTTCCGCTCGGCA 535
Db      464 CGGTATCTCCGAATCTCTCGGCGTGCAGAGCGGACCGCGCGCTTCCGCGTCTGA 523
Oy      536 CTGCGCGCGCGCTCGGCGTGGGTTTGTGCTCCCAAGTTCATGAGAGACCAAGACCTTG 595
Db      524 CGAGCGCTTCTGCTTCCGAGAGATCCGCGCGAGCGCAAGCG----- 568
Oy      596 TCGCGCTCGTCAACCGAGGAGCTCGCGCTGCTCATGACGCTCTGATGAGCGAGCGAGCA 655
Db      569 -----CATGCGCGAGATGAGCGGCTATCTTCCGCGCTCATGACCTCAAGCGCGGCG 622
Oy      656 ACCCGCTCGAAATGACGCTTGAAGATGCTGTTACGAGCGCGAGCGGACGAGCG 715
Db      623 AGGACGCGGAGAGACCTGCTCAGCGCGCTGTCGCGAGCAGAGCAGAGACGCGCTCCGCG 682
Oy      716 TTAGAGCAGAGAGCTGTCGCGCTGTCGAGGTCGATTTGCTGCTGTCGACCGATACCA 775
Db      683 TGAACCTCGAGAGCTGCTGCTGATGAGCCACATCTGCTGCTGCGGAGGACAGAACCA 742
Oy      776 CGATCACTTATACGCGCTGCTGCTGCTCAACCTGCTGCGGTCGCGCGAGCGCTGAGC 835
Db      743 CGGTCAATGATGTCGCAACCGGATGATAGCGCTGCTCTGCAACCGCGACAGCTGCGG 802
Oy      836 TGTGTAAGGCGGAGCGCGGCTCTGATGAGAAACGCGCTGATGAGAGTCTCCGCTTGACA 895
Db      803 CCTGCGGCGCGAGCATGACGCTTTCGAGCGCGGCTGAGAGAGATGTTGCGCTACGAG 862
Oy      896 AATATCTCAGATAGAAATGTCGCTTTCGCGGAGAGAGACCTGAGATGCTCGGCGCAT 955
Db      863 GCCCGGTGAATCCCGGACCTACCGCTTCCGCTGAGCGCGCTGACCTGAGAGGACGCG 922
Oy      956 CGATCAAGAAAGGAGATGTCCTTCTCTGATCCCGAGCGCGCTGAGAGATGAGGACTG 1015
Db      923 TCAATCCCGCGGTGACAGGTCCTGCTGCTGCTGCGGACCGCGCACCGCGCGCGAGC 982
Oy      1016 TATTTCTCAGGCGCAGCTGTTGATGTCGACGAGGACACGCGCGCGAGCTGCGTACG 1075
Db      983 GCTTCCCGGACCGGACCGCTTTCGACATCCGCGGAGACACCGCGCGCATCTTCGCTTG 1042
Oy      1076 GTAGAGGCGCGCATGTCGCTCCCGGGGTGCTTCTGCTGCGCTGAGAGGAGAGATGCGG 1135
Db      1043 GCGACGCGCATTCATCTTCTGATGCGCGCGCGCTTTCGCGGTTGAGAGCGCGGATGCGG 1102
Oy      1136 TGGGCGCATCTTCCGTAGGTTCCCGAGATGAGATGAAAGAAATCTCCG 1186
Db      1103 TCCGCGCGCTTCTCGAAGCTGCGCGGACCTGCGCTGAGAGCTCTCCCG 1153

```

RESULT 13
us-09-320-878-21

; Sequence 21, Application us/09320878A

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETHLACH, Melanie C.

; APPLICANT: BETHLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

```

; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320, 878A
; EARLIER FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119, 139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100, 880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087, 080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; us-09-320-878-21

```

Query Match 9.8%; Score 123; DB 3; Length 5970;

Best Local Similarity 47.5%; Pred. No. 3,3e-20;

Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;

```

Oy      236 AGTACTCTGCGGCACTTCCGAGCTCAGCGATATGAAAGTACGATGTTGTTGGGCTGC 295
Db      1582 ACTCCAGCAGTCCCTGACCGAGGCGGAGCGCGGCTCAACCAACATGCTGAGTCCG 1641
Oy      236 CGCCGAGAGATCAAGCTCGGGTCCGCAAGCTGTCACCCGCTTTTACATGCAAGCGGCA 355
Db      1642 ACCCGCGCGGACACCGCGGCTGCGCAAGCTGTCGCGCGCTGATGATGACATGCGCGG 1701
Oy      356 TCGACCTCTGCGCGCGCGGAAATACAGCGGACCGCTGACAGCTGCTGATGCTGCTCG 415
Db      1702 TCGAATCTCTGCGCGCGCGGCTGCAAGAGATGTCAGCGGCTGTCGAGCGCATGCTCG 1761
Oy      416 GACAAGAGAGTTCAGCTTGTGCGGATTAACGCGAGGGAATCCGATGCGCGGATCA 475
Db      1762 CGGCGCGCGGACCGCGCGCATCTGATGAGATCTCTGCGCGCGCTGCGCGCTGCG 1818
Oy      476 GCGCTCTGTGAAGTTCCGCGCGAGCTGTCAGAGAAATGTCGCTTCCGCTGCGGCA 535
Db      1819 CGGTATCTCCGACATCTCTGCGCTGCGCGGAGCGCGGACCGCGCGCTTCCGCTGGA 1878
Oy      536 CTGCGCGCGCGCTGCGCGTGGGTTTGTGCGCGGAGCTGAGTGAAGAGAGACCAAGACCTTG 595
Db      1879 CGAGCGCTTCTCTTCCGAGCATCCGCGCGGAGCGCGCGCGCGCGCGCGCGCG 1923
Oy      596 TCGCGCTCGTCAACGAGGAGCTCGCGCTGCTCATGAGCTCTGATGAGCGCGCAGAGA 655
Db      1924 -----CATGCGCGAGATGAGCGGCTATCTCTCCGCGTATGCACTTCAGAGCGCGGCG 1977
Oy      656 ACCCGGTGAATATACAGCTTTCGAGATGCTGCTTCAAGCGCGAGCGCGAGCGAGAGC 715
Db      1978 AGGACGCGGAGAGCTGTCAGCGCGCTGTCGCGGAGCAGACGACAGAGACGCGCTCCGCG 2037
Oy      716 TTAGAGCAGAGAGATGTCGCGCTGTCGAGGAGGATTAACGCTGTCGAGCAGATACCA 775
Db      2038 TGAACCTCGAGGAGTGTCTGATGAGCGCGCATCTGCTGCTGCGGCGGACGAGACCA 2097
Oy      776 CGATCACTTATCCGCTTCTGCTGCTCAACCTGCTGCGTGCAGCGCGGCTCGAGC 835
Db      2098 CGGTCAATGATTCGCAACGCGCATGATGCGGCTCTCTGCAACCGCGCATGCGCGG 2157
Oy      836 TGTGTAAGGCGGAGCGCGGCTCATGAGAGACGCGCTGATGAGTGTCTGCGCTTGACA 895
Db      2158 CCTCGCGGCGCATGAGCTCTTTCGAGCGCGGCTGAGAGAGATGTTGCGCTACGAGG 2217

```


Qy 896 ATATCTCAGATAGAACTGTGCTTTCCGAGCAGACCTGAGTACTGCGGGCAT 955
Db 2218 GCCCGGTGAATCCGACACTACCGCTTCCGGGTGAGCCGCTGACCTGAGCGACGG 2277
Qy 956 CGATCAGAAAGGGAGATGCTTTCTTCTGATCCGAGCGCCTGAGAGATGGAGACTG 1015
Db 2278 TCATCTCCGCGCGGTACACGGCTCTGCTGCTGCTGAGCGACGCCACCGCCGAGC 2337
Qy 1016 TATTTCCGAGCCAGACGCTGTTGATGTGCCAGCGGACACCGGCGCGAGCCTCGGTACG 1075
Db 2338 GCTTCCCGAGACCGACCGCTTCCGATCCGCGGAGACCGCGCCGACCTCTGCGCTTCG 2397
Qy 1076 GTAGAGCCCCCATCTGTGCCCCGAGGTGCTTCTGCTCGCCTCGAGCGAGATCGCG 1135
Db 2398 GCCACGGCATCACTTCTGACATCGCGCCCTTGGCGGGTGGAGGCGCGGATCGCG 2457
Qy 1136 TGGGACCATCTTCCGTAGCTTCCCGAGATGAACTGAAACTCCCG 1186
Db 2458 TCCGCGCCTTCTCGAACGCTGCGGACCTGCGCCCTGAGACGTCTCCCGC 2508

RESULT 14

US-09-141-908-11
; Sequence 11, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-11

Query Match 9.8%; Score 123; DB 4; Length 5970;
Best Local Similarity 47.5%; Pred. No. 3.3e-20;
Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;
Qy 236 AGTACTGTCGCGCATTCGCCGAGCTCAGCGATGATGAAAGTACGATTTGTCGGGCTGC 295
Db 1582 ACTCAGCATCTCCCTGACCGAGCGCGCGCTCAACACCAACATGCTGAGTCCG 1641
Qy 296 CGCCGAGATACACCTGCGGCTCCGCAAGCTGCTCAACCGGTGTTAGTACCGCGCA 355
Db 1642 ACCCGCGCGGACACCGCGCTGCGCAAGCTGCTGCGCGCGCGCGCGCGCG 1701
Qy 356 TCGACTCTGCGCGCGCGCAATACGCGACCGTGCACAGTGTCTGATGCTGCTCGC 415
Db 1702 TCGAGTCTGCGCGCGCGCGGTGCAAGATGCTGACGCGCTGCTGACGCGCATGTGG 1761
Qy 416 GACAGAGAGTTCAGCTGTGCGGAGTTACGCGGAGGATCCCGATGCGCGGATCA 475
Db 1762 CGCGCGCGCGCGCGCGCGATGATGAGATCTCCTGCGCGCGCGCGCGCGCG 1818
Qy 476 GCGCTCTGTGAAGGTTCCGCGCGAGTGTGACGAGAAATTCCGCTTCCGCTCGCGCA 535

Db 1819 CCGTATCTCCGAACCTCTCCGCGCTGCGCGAGCGGACCGCGCGCTTCCGCGCTCTGGA 1878
Qy 536 CTGCGCGCGCGCTGCGCGCTGCTTGTGATGCGCGAGTGTGATGAGAGACCAAGACCTCG 595
Db 1879 CCGAGCGCTTCTGCTTCCGAGCATCCGCGCGAGCGCGCGCGCGCGCGCGCG 1923
Qy 596 TCGGCTCCGATCAGGAGGAGCTGCGCTGCTCAATGACGTCTCGATGAGCGCGCGAGA 655
Db 1924 -----CATGCGCGAGATGAGCGGCTATCTCTCCGCGCTATGACCTCAAGCGCGG 1977
Qy 656 ACCGCTCGAAATGACCTTTGACATGCTGCTTCAAGCGGAGCGAGCGAGCGAGCG 715
Db 1978 AGGACGCGAGAGACCTGCTCAGCGCGCTGCGGAGACGAGAGAGAGAGCGCTCCG 2037
Qy 716 TGAGCAGAGAGAGCTGCTGCGCGCTCGGAGTATGCTGCTGCGCGAGCGAGTACCA 775
Db 2038 TGACCTCGAGAGAGTGTCTGCTGATGAGCGCGACATCTGCTGCTGCGCGAGCGAG 2097
Qy 776 CGATCTACCTTATCGCGCTGCTGCTCAACTGCTGCGCGCGCGCGCGCGCGCGCG 835
Db 2098 CGGTAACTGATGCGCGACGAGCATGATGCGCGCTGCTGCGACCGCGAGCGAGTGG 2157
Qy 836 TGTGAGAGCGCGAGCGCGCTCATGAGAAACGCGCTGATGAGTGTCTCGCTTCA 895
Db 2158 CCTGCGCGCGCGCATGACGCTTGTGACGCGCGCGGTGAGAGATGTTGCGCTGAG 2217
Qy 896 ATATCTCAGATAGAACTGTGCTTCCGAGCAGACCTGAGATGCTGCGGGCAT 955
Db 2218 GCCCGGTGAATCCGCGACCTACCGCTTCCGCTGAGACCGGTGACCTGAGCGGAC 2277
Qy 956 CGATCAGAAAGGGAGATGCTTCTTCTGATCCGAGCGCCTGAGATGGAGACTG 1015
Db 2278 TCATCTCCGCGCGGTGACAGCTTCTGCTGCTGCTGCGAGCGCGACCGCGCGAGC 2337
Qy 1016 TATTTCCGAGCCAGACGCTGTTGATGTGCCAGCGGACACCGCGCGAGCCTCGGTACG 1075
Db 2338 GCTTCCCGAGACCGACCGCTTCCGATCCGCGGAGACCGCGCCGACCTCTGCGCTTCG 2397
Qy 1076 GTAGAGCCCCCATCTGTGCCCCGAGGTGCTTCTGCTCGCCTCGAGCGAGATCGCG 1135
Db 2398 GCCACGGCATCACTTCTGACATCGCGCCCTTGGCGGGTGGAGGCGCGGATCGCG 2457
Qy 1136 TGGGACCATCTTCCGTAGCTTCCCGAGATGAACTGAAACTCCCG 1186
Db 2458 TCCGCGCCTTCTCGAACGCTGCGGACCTGCGCCCTGAGACGTCTCCCGC 2508

RESULT 15

US-09-657-440-21
; Sequence 21, Application US/09657440
; Patent No. 6509455
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT MABONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/657,440
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-657-440-21

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:29:26 ; Search time 308.318 Seconds
(without alignments)
11031.749 Million cell updates/sec

Title: US-10-014-717-1_COPY_62369_63628

Perfect score: 1 atgcacagagagcagcga.....agccctccaaagctgtag 1260

Sequence: 1 atgcacagagagcagcga.....agccctccaaagctgtag 1260

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1258.4	99.9	68750	21	AAZ55887
2	1246.2	98.9	71989	21	AAA29349
3	149.2	11.8	67459	25	ABQ77491
4	143.8	11.4	1257	22	AAAF81357
5	131.6	10.4	1400	12	AAQ14548
6	128.8	10.2	1347	25	ABX56047
7	123	9.8	5970	21	AAA75633
8	123	9.8	5970	21	AAZ56003

9	123	9.8	5970	25	ABZ56092
10	122.6	9.7	1191	24	ABN88910
11	121.4	9.6	1251	21	AAZ87301
12	120.8	9.6	1992	25	ABZ76672
13	119.6	9.5	16538	25	ABV75558
14	118.6	9.4	1449	24	ABN88911
15	116.8	9.3	1209	25	ABX56043
16	115	9.1	1879	12	AAQ11126
17	114.8	9.1	53789	19	AAV21187
18	114.6	9.1	1428	25	AAZ53024
19	114.4	9.1	1227	22	AAQ14511
20	114.4	9.1	1233	21	AAZ53730
21	114.4	9.1	12249	21	AAZ55840
22	114.4	9.1	15079	22	AAQ14499
23	114.4	9.1	18331	21	AAZ55857
24	112.8	9.0	15079	16	AAQ91580
25	111.8	8.9	1293	25	AAZ53020
26	111.8	8.9	1293	25	AAZ53026
27	111.4	8.8	1428	25	AAZ53027
28	111.4	8.8	1700	14	AAQ45569
29	110	8.7	966	24	ABK74899
30	109.6	8.7	1224	21	AAZ55788
31	109.4	8.7	103599	23	ABX04971
32	108.6	8.6	1293	25	AAZ53031
33	107.8	8.6	135638	25	ABX34289
34	107	8.5	1293	25	AAZ53022
35	107	8.5	2795	22	AAZ53022
36	107	8.5	2795	22	AAZ53022
37	106.6	8.5	1413	25	AAZ53019
38	106.4	8.4	2918	21	AAZ53922
39	106.4	8.4	2919	21	AAZ54466
40	98	7.8	1293	25	AAZ53052
41	97.4	7.7	50937	21	AAQ09469
42	95.4	7.6	1293	25	AAZ53028
43	94.8	7.5	1398	12	AAQ1127
44	94.6	7.5	47981	22	AAZ50757
45	94	7.5	1293	25	AAZ53032

ALIGNMENTS

AAZ55887	standard; DNA; 68750 BP.	
AAZ55887		
10-Apr-2000	(first entry)	
Sorangium cellulosum 68.75 kb contig.		
Epochlone biosynthesis; type I polyketide synthase; taxol substructure; anticancer; ds.		
Sorangium cellulosum.		
Location/Qualifiers		
1..1826		
/*tag= a		
/product= "Partial Orf 1 protein (AAZ55880)"		
/note= "No initiation codon given in the specification"		
complement (1900..3171)		
/*tag= b		
/product= "Orf 2 protein (AAZ55881)"		
3415..5556		
/*tag= c		
/product= "Orf 3 protein (AAZ55882)"		
complement (5612..5992)		
/*tag= d		
/product= "Orf 4 protein (AAZ55883)"		
6226..6675		

Contig 002 from co
Streptomyces sp. c
S. venezuelae macr
Streptomyces sp. T
Saccharopolyspora
Streptomyces sp. c
M. echinospora cal
Sequence encoding
Amycolatopsis medl
Streptomyces albol
Streptomyces clav
Hydroxylase orf3 1
Complete Mitomycin
Streptomyces clav
Complete nucleotid
S. clavuligerus cl
Streptomyces lydic
Streptomyces kaug
Streptomyces rimos
Sequence soyC and
Bacillus lichenifo
Hydroxylase orf4 1
S. cinamomensis m
Streptomyces lydic
S. atroolivaceus 1
Streptomyces chat
Streptomyces grise
Streptomyces grise
Streptomyces rimos
S. tendae nikomyc
Streptomyces tende
Streptomyces tuber
Streptococcus olea
Streptomyces tuber
Sequence encoding
Microspora meg
Streptomyces lydic

```

FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      7610..11875
CDS      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
CDS      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
CDS      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
CDS      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
CDS      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
CDS      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
CDS      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
CDS      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
CDS      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
CDS      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
CDS      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
CDS      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
CDS      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
CDS      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
CDS      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
CDS      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
CDS      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
FT
XX      WO996028-A2.
XX      23-DEC-1999.
XX      16-JUN-1999.
XX      99WO-EP04171.
XX      18-JUN-1998.
XX      98US-0099504.
XX      24-SEP-1998.
XX      98US-0101631.
XX      05-FEB-1999.
XX      99US-0118906.

```

```

PA      (NOVS ) NOVARTIS AG.
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX      Schupp T, Liqon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX      WPI: 2000-097741/08.
XX      P-Psdb: AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
XX      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
XX      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
XX      AAV58592, AAV58593, AAV58594.
XX      New isolated epoethione synthase genes, used for the recombinant
XX      production of epoethione for use in cancer therapy
XX      Claim 14, Page 87-104, 174pp; English.
XX
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-78) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiply drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
XX
XX      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
XX
XX      Query Match          99.9%; Score 1258.4; DB 21; Length 68750;
XX      Best Local Similarity 99.9%; Pred. No. 3.1e-279;
XX      Matches 1259; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX      1  ATGACACGAGGACGAGCGGATGAGTGAAGAGCGCTTGTCGACTTCAAGCCGTTG 60
XX      62369  ATGACACGAGGACGAGCGGATGAGTGAAGAGCGCTTGTCGACTTCAAGCCGTTG 62428
XX
XX      61  GCGCCTGGGTAACGCGAGGAGCCGTTCCCGGATGAGCGCTTGAGAGGCAACCCCC 120
XX      62429  GCGCCTGGGTAACGCGAGGAGCCGTTCCCGGATGAGCGCTTGAGAGGCAACCCCC 62488
XX
XX      121 ATCTTCTACTGGAGTGAAGCGCGCTCTCGGCTTCCACCCATACAGAGCTGCGCG 180
XX      62489  ATCTTCTACTGGAGTGAAGCGCGCTCTCGGCTTCCACCCATACAGAGCTGCGCG 62548
XX
XX      181 GGTTCGCGGAGGAGGAGCTGCGCGCTGAGTGAAGAGTGGGAATTCAGCGCGAGTAC 240
XX      62549  GGTTCGCGGAGGAGGAGCTGCGCGCTGAGTGAAGAGTGGGAATTCAGCGCGAGTAC 62608
XX
XX      241 TCGTCGGCCATTCGCGAGCTGAGCGATATGAAGAGTACGATTGTCGGGCTGCGCG 300
XX      62609  TCGTCGGCCATTCGCGAGCTGAGCGATATGAAGAGTACGATTGTCGGGCTGCGCG 62668
XX
XX      301 GAGGATCACGCTCGGCTGCGCAAGCTGCTCAACCCGTCGTTACGTCACGCGCCATCAC 360
XX      62669  GAGGATCACGCTCGGCTGCGCAAGCTGCTCAACCCGTCGTTACGTCACGCGCCATCAC 62728
XX
XX      361 CTGCTGCGGCGCGAATACAGGCAACGTCGACGAGCTGCTGATGTCGCTCCGAGCA 420

```

Db	62129	CTGTCGCGCCCGAATAACACCGCACCGTGCACCAAGTCTGCTCATGCTGCTCCGAGCAA	62788
Qy	421	GAGAGATTTCGACGTTGTGCGGGAATTACGCGAGGGAATCCGATGCGCGATCAAGCT	480
Db	62789	GAGAGATTTCGACGTTGTGCGGGAATTACGCGAGGGAATCCGATGCGCGATCAAGCT	62848
Qy	481	CTGTTGAAGGTTCCGCGCGCACTGTGACGAAAGTTCCGTGCTTCGCGTCCGCGACTGG	540
Db	62849	CTGTTGAAGGTTCCGCGCGCACTGTGACGAAAGTTCCGTGCTTCGCGTCCGCGACTGG	62908
Qy	541	CGCGCGCTCGGCGGTGGGTTGGTGGTCCCAAGTGCATGAGAGACAAAGCCCTGGTGGG	600
Db	62909	CGCGCGCTCGGCGGTGGGTTGGTGGTCCCAAGTGCATGAGAGACAAAGCCCTGGTGGG	62968
Qy	601	TCCGTCAACGAGGGGCTCGCGCTGCTCATGACCTCTCGATGAGCGCGCAGAAACCG	660
Db	62969	TCCGTCAACGAGGGGCTCGCGCTGCTCATGACCTCTCGATGAGCGCGCAGAAACCG	63028
Qy	661	CTCGAATAATGACCTTCTGACGATGTGCTTCACGGCGAGCCGACGGCAGAGCTGAGC	720
Db	63029	CTCGAATAATGACCTTCTGACGATGTGCTTCACGGCGAGCCGACGGCAGAGCTGAGC	63088
Qy	721	ACGAAAGAGCTGTGTCGCGCTCGTGGGTCGATTAATCGTCTGTGACACCGATACAGATC	780
Db	63089	ACGAAAGAGCTGTGTCGCGCTCGTGGGTCGATTAATCGTCTGTGACACCGATACAGATC	63148
Qy	781	TACCTTATCGCGTTGCGTGTGCTCAACCTGCTCGGTCGCGCGAGCGCTCGAGCTGTG	840
Db	63149	TACCTTATCGCGTTGCGTGTGCTCAACCTGCTCGGTCGCGCGAGCGCTCGAGCTGTG	63208
Qy	841	AAGGCCGAGCCCGGGGCTCATAGAGAAACGCGCTGATGAGGTGCTTCGCTTCGAAATATC	900
Db	63209	AAGGCCGAGCCCGGGGCTCATAGAGAAACGCGCTGATGAGGTGCTTCGCTTCGAAATATC	63268
Qy	901	CTCGAATAGGAACCTGTGCGCTTCGCGCAGGACGACCTGGAATGACTCGGGGATGATC	960
Db	63269	CTCGAATAGGAACCTGTGCGCTTCGCGCAGGACGACCTGGAATGACTCGGGGATGATC	63328
Qy	961	AAGAAAGGGAGATGATCTTCTTCTCTGATCCGAGCGCCCTGAGAGATGGACTGTATTC	1020
Db	63329	AAGAAAGGGAGATGATCTTCTTCTCTGATCCGAGCGCCCTGAGAGATGGACTGTATTC	63388
Qy	1021	TTCAGGGCCAAACGCTTGTGATGTGCGACGGGACACGGGCGCGAGCTCCGCTACGGTAA	1080
Db	63389	TTCAGGGCCAAACGCTTGTGATGTGCGACGGGACACGGGCGCGAGCTCCGCTACGGTAA	63448
Qy	1081	GGCCCCCATGTCTGCCCCCGGGGTGTCCCTTGTCTCGCTCGAGGCGAGATGCCGTGGGC	1140
Db	63449	GGCCCCCATGTCTGCCCCCGGGGTGTCCCTTGTCTCGCTCGAGGCGAGATGCCGTGGGC	63508
Qy	1141	ACCATCTTCCGTAGGTTCCCGAGATGAAGCTGAAAGAACTCCGCTGTTGGATACAC	1200
Db	63509	ACCATCTTCCGTAGGTTCCCGAGATGAAGCTGAAAGAACTCCGCTGTTGGATACAC	63568
Qy	1201	CCCGCGTTCCGGAAACATCGAATCACTTAAGCTCATCTTGAAAGCCCTTCAAAGTGTGATG	1260
Db	63569	CCCGCGTTCCGGAAACATCGAATCACTTAAGCTCATCTTGAAAGCCCTTCAAAGTGTGATG	63628

RESULT 2	
AAA29349	
ID	AAA29349 standard; DNA, 71989 BP.
XX	
AC	
XX	AAA29349;
DT	
DT	12-SEP-2000 (first entry)
XX	
XX	
DE	Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX	
KW	Epothilone; polyketide synthase; epoa; epob; epoc; epod; epoe; epof; epog; epok; epok; P450 epoxidase; ORFa; ORFb; promoter; enhancer; anti-fungal; tubulin polymerization assay; anti-tumour; cytostatic; ds.

XX	OS	Location/Qualifiers
XX	Sorangium cellulosum.	
FH	Key	3..992
FT	CDS	/*tag= a
FT		/label= ORF_A
FT		/product= transposase
FT		/note= "not part of the PKS"
FT		989..1501
FT	CDS	/*tag= b
FT		/label= ORF_B
FT		/product= transposase
FT		/note= "not part of the PKS"
FT		1998..6263
FT	CDS	/*tag= c
FT		/label= epa gene
FT		/note= "encodes the loading domain"
FT		2031..3548
FT		/*tag= d
FT		/note= "encodes ketide synthase (KS-Y) of the loading domain"
FT		3621..4661
FT	misc_RNA	/*tag= e
FT		/note= "encodes acyl transferase (AT) of the loading domain"
FT		4917..5810
FT	misc_RNA	/*tag= f
FT		/note= "encodes enoyl reductase (ER) of the loading domain, potentially involved in formation of the thiazole moiety"
FT		5856..6155
FT	misc_RNA	/*tag= g
FT		/note= "encodes acyl carrier protein (ACP) of the loading domain"
FT		6260..10493
FT	CDS	/*tag= h
FT		/label= epob_gene
FT		/note= "encodes module 1, the NRPS module"
FT		2031..3548
FT	misc_RNA	/*tag= i
FT		/note= "encodes condensation domain C2 of the NRPS module"
FT		2031..3548
FT	misc_RNA	/*tag= j
FT		/note= "encodes condensation domain C2 of the NRPS module"
FT		6861..6887
FT	misc_RNA	/*tag= k
FT		/note= "encodes heterocyclization signature sequence"
FT		6861..6887
FT	misc_RNA	/*tag= l
FT		/note= "encodes condensation domain C4 of the NRPS module"
FT		7358..7366
FT	misc_RNA	/*tag= m
FT		/note= "encodes condensation domain C7 (partial) of the NRPS module"
FT		7898..7921
FT	misc_RNA	/*tag= n
FT		/note= "encodes adenylation domain A1 of the NRPS module"
FT		7898..7921
FT	misc_RNA	/*tag= o
FT		/note= "encodes adenylation domain A1 of the NRPS module"
FT		8261..8308
FT	misc_RNA	/*tag= p
FT		/note= "encodes adenylation domain A3 of the NRPS module"
FT		8411..8422
FT	misc_RNA	/*tag= q
FT		/note= "encodes adenylation domain A4 of the NRPS module"
FT		8861..8905
FT	misc_RNA	/*tag= r
FT		/note= "encodes adenylation domain A6 of the NRPS module"

```
FT misc_RNA 8966..8983
FT /tag= b
FT /note= "encodes adenylation domain A7 of the NRPS module"
FT misc_RNA 9090..9179
FT /tag= t
FT /note= "encodes adenylation domain A8 of the NRPS module"
FT misc_RNA 9183..9992
FT /tag= u
FT /note= "encodes oxidation region for forming thiazole"
FT misc_RNA 10121..10138
FT /tag= v
FT /note= "encodes adenylation domain A10 of the NRPS module"
FT misc_RNA 10261..10306
FT /tag= w
FT /note= "encodes thiolation domain (PCP) of the NRPS module"
FT CDS 10639..16137
FT /tag= x
FT /label= epoc_gene
FT /note= "encodes module 2"
FT misc_RNA 10654..12033
FT /tag= y
FT /note= "encodes KS2, the KS domain of module 2"
FT misc_RNA 12250..13287
FT /tag= z
FT /note= "encodes AT2, the AT domain of module 2"
FT misc_RNA 13327..13899
FT /tag= aa
FT /note= "encodes dehydratase (DH) 2, the DH domain of module 2"
FT misc_RNA 14962..15756
FT /tag= ab
FT /note= "encodes ketoreductase (KR) 2, the KR domain of module 2"
FT misc_RNA 15763..16008
FT /tag= ac
FT /note= "encodes ACP2, the ACP domain of module 2"
FT CDS 16134..37907
FT /tag= ad
FT /label= epod_gene
FT /note= "encodes modules 3-6"
FT misc_RNA 16425..17606
FT /tag= ae
FT /note= "encodes KS3"
FT misc_RNA 17817..18857
FT /tag= af
FT /note= "encodes AT3"
FT misc_RNA 19581..20396
FT /tag= ag
FT /note= "encodes KR3"
FT misc_RNA 20424..20642
FT /tag= ah
FT /note= "encodes ACP3"
FT misc_RNA 20706..22082
FT /tag= ai
FT /note= "encodes KS4"
FT misc_RNA 22296..23336
FT /tag= aj
FT /note= "encodes AT4"
FT misc_RNA 24069..24647
FT /tag= ak
FT /note= "encodes KR4"
FT misc_RNA 24867..25151
FT /tag= al
FT /note= "encodes ACP4"
FT misc_RNA 25203..26576
FT /tag= am
FT /note= "encodes KS5"
FT misc_RNA 26793..27883
FT /tag= an
FT /note= "encodes AT5"
FT misc_RNA 27966..28574
```

```
FT /tag= ao
FT /note= "encodes DH5"
FT misc_RNA 29433..30287
FT /tag= ap
FT /note= "encodes ER5"
FT misc_RNA 30321..30869
FT /tag= aq
FT /note= "encodes KR5"
FT misc_RNA 31077..31373
FT /tag= ar
FT /note= "encodes ACP5"
FT misc_RNA 31440..32807
FT /tag= as
FT /note= "encodes KS6"
FT misc_RNA 33018..34067
FT /tag= at
FT /note= "encodes AT6"
FT misc_RNA 34107..34676
FT /tag= au
FT /note= "encodes DH6"
FT misc_RNA 35760..36641
FT /tag= av
FT /note= "encodes ER6"
FT misc_RNA 36705..37256
FT /tag= aw
FT /note= "encodes KR6"
FT misc_RNA 37470..37769
FT /tag= ax
FT /note= "encodes ACP6"
FT CDS 37912..49308
FT /tag= ay
FT /label= epod_gene
FT /note= "encodes modules 7 and 8"
FT misc_RNA 38014..39375
FT /tag= az
FT /note= "encodes KS7"
FT misc_RNA 39589..40626
FT /tag= ba
FT /note= "encodes AT7"
FT misc_RNA 41341..41922
FT /tag= bb
FT /note= "encodes KR7"
FT misc_RNA 42181..42423

Query Match 98.9%; Score 1246.2; DB 21; Length 71989;
Best Local Similarity 99.4%; Pred. No. 2e-276;
Matches 1251; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 ATGACACAGGACGACGATCAGATGAGACGAGCCGCTTTCGACTTCAAGCCGTTTC 60
DB 56757 ATGACACAGGACGACGATCAGATGAGACGAGCCGCTTTCGACTTCAAGCCGTTTC 56816
QY 61 GCGCCTGGGTACGCGAGAGACCCGTTCCCGCGATCGAGCGCTGAGAGGCAACCCCC 120
DB 56817 GCGCCTGGGTACGCGAGAGACCCGTTCCCGCGATCGAGCGCTGAGAGGCAACCCCC 56876
QY 121 ATCTTCTACTGGAGATGAAAGCCGCTCCGCGCTCCACCCGATACACAGAGTGCAGCG 180
DB 56877 ATCTTCTACTGGAGATGAAAGCCGCTCCGCGCTCCACCCGATACACAGAGTGCAGCG 56936
QY 181 GTGTCGCGAGACGCTTCGCGGTCACTAGTCAGAGAGAGTGGAAATCGAGCCGAGTAC 240
DB 56937 GTGTCGCGAGACGCTTCGCGGTCACTAGTCAGAGAGAGTGGAAATCGAGCCGAGTAC 56996
QY 241 TCGTCGCGCATTCGCGAGCTCAGCGATATGAAGAGTACGATTTGCGGCTGCGCGCG 300
DB 56997 TCGTCGCGCATTCGCGAGCTCAGCGATATGAAGAGTACGAGTGTTCGCGCTGCGCGCG 57056
QY 301 GAGGATCACGCTCGGCTCGCAAGCTCGTCAACCCGCTGTTTACGTCAAGCCCATTCAGC 360
DB 57057 GAGGATCACGCTCGGCTCGCAAGCTCGTCAACCCATGTTTACGTCAAGCCCATTCAGC 57116
QY 361 CTGCTGCGCGCGGAAATACAGGCAACCGTCAGTCAGTGTCTGCTCGGAGCA 420
```



```

Db      57117 CTGTCGCCGCCGAAATCAGGCGCACCGTCGACGCTGCTCGATGCTCGCTCCGAGCAA 57176
Oy      421 GAGAGATTGACGATTGTCGGGATTACGCGGAGGGAATCCGATGCGGCGATCGAGCCT 480
Db      57177 GAGAGATTGACGATTGTCGGGATTACGCGGAGGGAATCCGATGCGGCGATCGAGCCT 57236
Oy      481 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAATTCCGTCGCTTCGGCTCGGCGACTGCG 540
Db      57237 CTGTTGAAGGTTCCGGCCGAGTGTGACGAGAAATTCCGTCGCTTCGGCTCGGCGACTGCG 57296
Oy      541 CGCCGCGCTCGCGCTGTTGGTGTGCCAGGTTCATGAGAGACCAAGACCTGCTGCGCG 600
Db      57297 CGCCGCGCTCGCGCTGTTGGTGTGCCAGGTTCATGAGAGACCAAGACCTGCTGCGCG 57356
Oy      601 TCCGTCACCGAGGGGCTGCGCGCTGCTCATGACGTCCTCGATGAGCGGCGGAGAACCCG 660
Db      57357 TCCGTCACCGAGGGGCTGCGCGCTGCTCATGAGCGTCCTCGATGAGCGGCGGAGAACCCG 57416
Oy      661 CTCGAAATGACGTCCTTACGATGCTGCTCAGCGCCGAGCCGACGAGAGGCTGAGC 720
Db      57417 CTCGAAATGACGTCCTTACGATGCTGCTCAGCGCCGAGCCGACGAGAGGCTGAGC 57476
Oy      721 ACGAAGAGCTGCTGCGCTGCTGAGGTCGATTAATGCTGCTGCGACCGATACACGATC 780
Db      57477 ACGAAGAGCTGCTGCGCTGCTGAGGTCGATTAATGCTGCTGCGACCGATACACGATC 57536
Oy      781 TACCTTATCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db      57537 TACCTTATCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57596
Oy      841 AAGGCGGAGCCCGGCTCATGAGAACCGGCTCATGAGTGTCTCCGCTTCGACAAATATC 900
Db      57597 AAGGCGGAGCCCGGCTCATGAGAACCGGCTCATGAGTGTCTCCGCTTCGACAAATATC 57656
Oy      901 CTCAGATAGGAATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db      57657 CTCAGATAGGAATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 57716
Oy      961 AAGAAAGGAGATGTCCTTCTCTGATCCCGAGGCGCTGAGAGATGGAGATCTATATC 1020
Db      57717 AAGAAAGGAGATGTCCTTCTCTGATCCCGAGGCGCTGAGAGATGGAGATCTATATC 57776
Oy      1021 TCCAGGCGAGAGCTGTTGATGTCGAGCGGACACGCGGCGGAGCTGCGCTACGCTAGA 1080
Db      57777 TCCAGGCGAGAGCTGTTGATGTCGAGCGGACACGCGGCGGAGCTGCGCTACGCTAGA 57836
Oy      1081 GGGCCCCATGTCGCCCCGGGGGTGTCCTTGTCTGCTGCTGAGGCGGAGATGCGCGTGAGC 1140
Db      57837 GGGCCCCATGTCGCCCCGGGGGTGTCCTTGTCTGCTGCTGAGGCGGAGATGCGCGTGAGC 57896
Oy      1141 ACCATCTTCCTAGATTCCCGAGATGAGAGTGAAGAAATCTCCGCTGTTGATACAC 1200
Db      57897 ACCATCTTCCTAGATTCCCGAGATGAGAGTGAAGAAATCTCCGCTGTTGATACAC 57956
Oy      1201 CCCGCGTCCGGAACATGCAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 1259
Db      57957 CCCGCGTCCGGAACATGCAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGGATA 58015

```

RESULT 3
ABQ77491/c
ID - ABQ77491 standard; DNA: 67459 BP.

XX ABQ77491;

XX 14-MAY-2003 (first entry)

DE S. aurentiaca DNA containing sti gene cluster.

KM stiA; stiB; stiC; stiD; stiE; stiF; stiG; stiH; stiJ; stiK; stiL;
KW aromatic structure biosynthesis; beta-ketoacyl synthase;
XX acyltransferase; dehydratase; enoylreductase; beta-ketoacylreductase;

```

KW acyl carrier protein; O-methyl transferase; ds.
XX
XX Stigmatella aurentiaca.
XX
XX Location/Qualifiers
XX Key 14345..21466
XX CDS /*tag= a
FT /product= "stiA"
FT /note= "GTG start codon"
FT 21463..26220
FT /*tag= b
FT /product= "stiB"
FT 26284..31941
FT /*tag= c
FT /product= "stiC"
FT 31934..37738
FT /*tag= d
FT /product= "stiD"
FT 37740..43553
FT /*tag= e
FT /product= "stiE"
FT 43546..50202
FT /*tag= f
FT /product= "stiF"
FT 50205..54401
FT /*tag= g
FT /product= "stiG"
FT 54398..59212
FT /*tag= h
FT /product= "stiH"
FT 59223..63002
FT /*tag= i
FT /product= "stiJ"
FT 63070..63840
FT /*tag= j
FT /product= "stiK"
FT 63929..65470
FT /*tag= k
FT /product= "stiL"
FT /note= "GTG start codon"
FT
FT DE10128661-A1.
XX
XX 19-DEC-2002.
XX
XX 15-JUN-2001; 2001DE-1028661.
XX
XX 15-JUN-2001; 2001DE-1028661.
XX
XX (GBF) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX
XX Kunze B, Nordstiek G, Siakowski B, Bioecker H, Galtatzis N;
XX Mueller R;
XX WPI; 2003-185404/19.
XX
XX Nucleic acid molecule comprising one or more genes from Stigmatella
XX species -
XX
XX Claim 3; Page 9-43; 48pp; German.
XX
XX This invention describes a novel nucleic acid molecule comprising one or
XX more genes from Stigmatella species selected from stiA, stiB, stiC, stiD,
XX stiE, stiF, stiG, stiH, stiJ, stiK and stiL. The nucleic acid is used for
XX biosynthetic purposes, particularly for producing aromatic structures.
XX The products of the invention have domains sharing high homology with
XX beta-ketoacyl synthase (KS), acyltransferase (AT), dehydratase (DH),
XX enoylreductase (ER), beta-ketoacylreductase (KR), acyl carrier protein
XX (ACP), O-methyl transferase (O-MT). This sequence represents a
XX polynucleotide isolated from S. aurentiaca which contains the sti gene
XX cluster described in the disclosure of the invention.
XX
XX Sequence 67459 BP; 10755 A; 20921 C; 23348 G; 12435 T; 0 other;
XX

```

Query Match	Similarity	11.8%	Score 149.2	DB 25	Length 67459
Best Local	Similarity	49.2%	Pred. No. 1.7e-24		
Matches	Conservative	0	Mismatches 458	Indels	Gaps 2
Qy	298	CCGAGAGATCAACGCTCGGGTCCGACAGCTCGTCAACCCGCTGTTACGTCAACGCCCATC	357		
Db	5928	CCGCGCGTGCACACGCGGGGTGGCCGACTGTATCGTGGGCGCGCTGACGCCCGCGCGTTG	5869		
Qy	358	GACCTGCTCCGCGCCGAAATACAGGGCACCGTCCACACGCTGCTGTGATGCTCGCTCCGGA	417		
Db	5868	TCGACAAATGAGACCCAGGGTTGGGACCCCTGTGTGACCGGCTGTGAGAGGGCTGGCCGTG	5809		
Qy	418	CAAGAGAGATTGACGTTGTGCGGGATTACGCGAGGGGAATCCCGATCCGCGCATCAGC	477		
Db	5808	AAAGGCGCGGACAGCTCATGAGGACTTCGCGCCCGCCCATCCCATGAGATTAATTTGCG	5749		
Qy	478	GCTCTGTTGAAGTTCCGCGCGAGTGTGACGAGAAATTCGTCGCTTCGCTCGCGACT	537		
Db	5748	AACTGCTGTGACATCCCGACGGAGGAGCGGGGCGCGTGGCGGGGCTGTGCTCGCATTC	5689		
Qy	538	GCGCGCGCGCTTCGCGGTGGGTTTGTGTCGCCCAAGTCAATGAGAGAACCAACCTGCTG	597		
Db	5688	CTCGCGCGGCTCGAACCCCGTCTGACGCGCGAGGAGGCGCTCGCGGAAATGAGCGGTA	5629		
Qy	598	GCGTCCGTACCGAGGGGCTCGCGCTGCTCCATGACGTCCTCGATGACCGCGCGAGAAC	657		
Db	5628	ACAGAGTTCTGTGACTACTCCGACTCTTCGTGCGGACGCGCGAGCGCGCCGGGAGAC	5568		
Qy	658	CCGCTCGAAATGACGTCCTTGACGATGCTGCTTAGCGCCGAGGCGGACGCGACGAGCTG	717		
Db	5568	CC---CGCAACGAGAGTGTCTGACGCGGCTCATCCAGGGCGAATCCGAGGGAAAGATTG	5512		
Qy	718	AGCAGAGAGAGCTGTGCGCTGCGCTGTGGGTGCAATATCGCTGCTGGACCGATACAGC	777		
Db	5511	ACCGAGAGAGAGCTGTGCTGACACAGTGCCTTTCTCTGAACCGCGGGCACGAGACGACA	5452		
Qy	778	ATCTACCTTAATCGACGTTGCTGTGCTCAACCTGCTGCGAGTGGCCCGAGGCGCTCGACGTG	837		
Db	5451	ACGAACTCATGTGGAGACGCGCTGGAACTGCTGGCGGCTTCCGGAAGACGCGGCTCGG	5392		
Qy	838	GTAAGGCGGACGCCGCGGCTCATGAGGAACGCGCTCGATGAGAGTCTCCGCTTGACAAAT	897		
Db	5391	CTGCTGCAGGCCCGCGCTGTATCCCGACGGGAGTGAAGAGTCTCTCGTTACGAAATCC	5332		
Qy	898	ATCCTCAGAAATAGAACTGTGCGTTTCCGACGAGGACCTGAGATACGTGGGGCGATCG	957		
Db	5331	TCCAACCACTCGGAGACCGCGGGGTGCG--GGAGGACAGGAAATGGCGGCGCTAACG	5279		
Qy	958	ATCAAGAAAGGGAGATGATGCTTTTCTCCGTAATCCGACCGGCTTGAGATGGGACTTA	1017		
Db	5274	GTCGCGACCGGACCTTCTCAACGCTCTGCAATCGGCGCGCGCCACACCGGACCCCGCGGC	5215		
Qy	1018	TTCTTCACGACGACAGCTGTGATGTGACGCGGACACGCGCGCGAGCTTCGCTGATCGGT	1077		
Db	5214	TTTCGAGAACCCAGACCACTTCGACCTGGGCGGCCAGGCCAACCCGCACTTCGCTTCGCG	5155		
Qy	1078	AGAGGCGCCCAATGTGCGCCCGGGGTGTCCCTGTGCTCGCTCGAGGGCGAGATGCGCGTG	1137		
Db	5154	GCGGGGGCGCAATCGTGGCGGGAGTGAACCTTGTGCGGATGAGAGCGCGGATACGTTTG	5095		
Qy	1138	GGCACCACTTCTCCGTAGGTTCCCGCAGATGGAAGCTGAAGAAACTCCGCTTTGGATAC	1197		
Db	5094	GCGGCGTTCTGTGCGCGCTTCCCGGACATATGCGTACCGGACCGCGCGGTGAGGGCGGCG	5035		
Qy	1198	CACCCCGCGTTCCG 1211			
Db	5034	CGGCGCGCGATTCCG 5021			

Query Match	Best Local Similarity	Matches	Score	DB	Length
11.4%	48.7%	0	143.8	1257	
455	Conservative	0	Mismatches 447	Indels 33	Gaps 1
254	CCGAGCTCAGCGATGAGAAAGTACGAGATTGTCGGGCTGCCGCCGAGAGATCACGCTC	313			
233	CGGGAGATCGGGAAGACTGACCAACATGCTCAACCTCGAACCCCGGACCATATCC	292			
314	GGGTCGCCAGCTGTCAACCCGTCGTTTACGTACGCGCGCATGTGACTCTGGCGCCG	373			
293	GCCGCGCTCTGTTGGCGCGGTTTCAACCCGCGCCAGGTGAGCGCTGTCAACCGC	352			
374	AAATACAGCGCAGCGTGCACCGATGCTGCATGCTCCCTCGGACAAGAGATTGCAGC	433			
353	ATATGACGATACCGAGCATTTGCTGACGCGCATATGCGCGCGGGAACAGCGCAC	412			
434	TTGTGCGGATTAACGCGAGGAAATCCGATGCGCGCATGACGCTCTGTGAAGTTT	493			
413	TGATGCGCGCATTTGCGCATCCCGCTGCACATCGCGGTGATCTTCGACGCTGTGGCATTC	472			
494	CGGCGCATGTGACGAGAAAGTTCCGTCGCTCGGCTGGGAGACTGCGCGCGCTCGGGC	553			
473	CGAGAGCGGAGCGGAACACGCGCCGCAATCTT-----	506			

```

QY 554 TGGGTTTGGTGGCCCGAGTGCATGAGAGACCAAGACCTGTCGCTCCGTCACCCGAGG 613
DB 507 -----GGAGGCGCAGGCGGGAACGTGTGCGCCGAGAGAGGCCACAGGCGCTGGCGAATG 559
QY 614 GGCCTGCGCTGCTCCATGACGTCTCTGATGAGCGGCGCAGAACCCGCTCGAAAAATGACG 673
DB 560 CGCAGGTGCACTACCTGCGCGCTGTCTGTCGAGGCCAAGGCGCGGACGCGCGCCGACGACG 619
QY 674 TCTTGAACATGCTGCTTCAAGCGCGGAGCGCGACAGAGCTGAGCAAGAAAGAGCTG 733
DB 620 TCTACAGCGCGGTGTGTGACGCGCGCGCGACGAGAGCGCGCATGTGACGAAAGCGAACTCG 679
QY 724 TCGCGCTGCGGAGTGCATATATGCTGTGACGACCGATACAGATTAATCTTATCGCGT 793
DB 680 TCTCATGAGCCCACTGCTGATGATGAGAGCGCTTGAAGACCAATGAACATGATCGGCA 739
QY 794 TCGCTGTCTCAACTGCTGCGGTCGCCGAGCGCGCTGAGCTGTGAAGCGGACCGCG 853
DB 740 ACGCGCTGTCACTGCTGTGTCAACCCGAGCAACTGGCGTGTGCTGCGCGCGCAGCGG 799
QY 854 GGCCTATGAGAACCGCTCGATGAGTGTCTCGCTTCAATATCTCAAGAAATAGAA 913
DB 800 AACTCTGCGCAACCGCATGAGAACTGTGTCCGACAGACGTCGCGGTGCGCGCTCGA 859
QY 914 CTGTGCTGTTTCCGACGAGCAGCATGATGCTGCGGCGCATGATCAAGAAAGGAGAA 973
DB 860 TGTGCGCTTCACTGCGTGAAGACGTGAACTGAGACGGGTCCATTCGCCGCGCGCAT 919
QY 974 TGTGCTTCTCTGATTCGAGCGCGCTGTGAGATGGAATGTATTTCTCAGAGCCAGACG 1033
DB 920 ACATCTGCTGCTTCAACCTGACCGCAACCAAGATGCGAGCGCTTCAAGATCCCGAC 979
QY 1034 TGTTGAATGTGAGAGGAGCAACGGGCGCGAGCTGCGGATGAGAGGCGCGCATGTCT 1093
DB 980 GCTTGACCTCAACCGCAACACGATGCGCATCTGCGCTTCCGCTGCGCTGACT 1039
QY 1094 GCCCGGGGGTGTCCCTGCTGCGCTGAGAGCGAGATGCGCGTGGGACATCTTCCGTA 1153
DB 1040 GCGTGGGCGCTGCTGCGCGCGCTGAGAGGCGGATGCCCATCCAGCGCTGCTCGCG 1099
QY 1154 GGTTCGCCGAGATGAAGCTGAAAGAACTCCGCTG 1188
DB 1100 GCTTCCCGACCTCCAGTTGGGCGGTGCCCAAGCG 1134

```

RESULT 5
AA014548 standard; DNA; 1400 BP.

```

XX AC AA014548;
XX DT 28-JAN-1992 (first entry)
XX DE EryF gene.
XX KM C-6 hydroxylation: erythromycin; 6-deoxyerythromycin; antibiotics;
XX KW saccharopolyspora; cytochrome P450 monooxygenase; ss.
XX OS Saccharopolyspora erythraea.
XX FH Key location/Qualifiers
XX FT CDS 158..1372
XX ET /tag= a
XX PN WO9116334-A.
XX PD 31-OCT-1991.
XX PF 16-APR-1991; 91WO-US02600.
XX PR 18-APR-1990; 90US-0510483.
XX

```

PA (ABBO) ABBOTT LABORATORIES.

XX Weber JM;

XX WPI; 1991-339744/46.

XX P-PSDB; AARI4724.

XX New 6-deoxy:erythromycin deriva. - are antibiotics with increased

XX acid stability, produced by cultivation of saccharopolyspora.

XX Disclosure; Fig 3; 56pp; English.

XX The eryF gene encodes the 6-hydroxylase component of the cytochrome

XX P450 monooxygenase system responsible for the hydroxylation of 6-de-

XX oxyerythronolide B to erythronolide B. Interruption of this step

XX results in the formation of deoxyerythromycin A and new derivatives

XX useful as antibiotic which have better stability against acids than

XX the corresponding erythromycins. Interruption of the reaction can

XX be effected by an insertion into the eryF gene of a plasmid, gene

XX replacement or chemical or light-induced mutagenesis.

XX The gene is positioned between the eryH and eryG genes.

XX See also AA014549.

XX Sequence 1400 BP; 212 A; 495 C; 479 G; 214 T; 0 other;

XX Query Match 10.4%; Score 131.6; DB 12; Length 1400;

XX Best Local Similarity 49.1%; Pred. No. 9.5e-21;

XX Matches 430; Conservative 0; Mismatches 419; Indels 27; Gaps 2;

QY 298 CCGAGAGATCAAGCTCGGCTCGCAAGCTGTCAACCCGCTGTTACGTACGCGCATC 357

DB 440 CCGCGACCCACACCCGGCTCGCAAGCTGTGTGTCAGAGATTCAACCGCTCGCGTG 499

QY 358 GACCTGTGCGCGCGCAATATACAGCGCACCGTCAACCGCTGTGATGCTGCTCCGGA 417

DB 500 GAGCGATGCGCGCGCGCGTGTGAGATCAACCGCGAGCTGTGAGAGAGTGGCGAC 559

QY 418 CAAGAGAGTTTCAAGTTGTGCGGATTAACCGAGAGAAATCCGATCGCGCATACG 477

DB 560 TCGCGCTGTGTGATGTATGTATACCGCTTCCGCCACCGCGTCCATCAAGTATCTGC 619

QY 478 GCTCTGTAAGATTCCGCGCGAGTGTGAGAGAAATTCGCTGCTGCGCTCGCGACT 537

DB 620 GAGCTGCTCGCGCTGTGAGAGAAATCCGCGAGGATTCGAGCGGTGAGCTCGGAGA- 677

QY 538 GCGCGCGCTCGCGCGTGTGTGCTGCCAGGTGTATGAGAGAACCAACCTGTGTC 597

DB 678 -----TCTGTGTATGTGACCCGAGCGCGGCGGAACAGCGCGG 715

QY 598 GCGTCCGTACAGAGGCGCTGCGCTGCTCCATGACGCTCTGATGACGCGGCGAGAAC 657

DB 716 CAGCGCGCAGAGAGGTGTCAACTTCACTTCAAGCTGTGTGAGCGCGCGCGACGAG 775

QY 658 CCGCTCGAAATGACGTGTGTGACGATGCTGTTCAGCGCGAGCGCGACGAGCGTGTG 717

DB 776 CCGCGCGACGACGTGTGTGCGCGCTTATCAAGGCTCAGAGAGAGATGACGAGTGTG 835

QY 718 AGCAGAGAGAGCTGTGTGCGCTGTGTGTGATATGCTGTGTGACGAGTATCCAG 777

DB 836 AGCGCGAGAGCTGT 895

QY 778 ATTTACCTTATCGGCTGT 837

DB 896 GTGAGCTTATCGGATGT 955

QY 838 GTGAAGCGGAGCGCGGCTCATGAGAACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 897

DB 956 GTGCGCGGAGACCGGT 1015

QY 898 ATCTCAAGATGAGAACTGT 957

DB 1016 CTGACGAGAGAC---CACACGCGCTTGT 1072

QY 958 ATCAAGAAAGGAGATGCTTTCTTCCTGATCCGAGGCCCTGTAGATGAGACTGA 1017
DB 1073 ATCCCCCAATACAGACGGGTGCTGTGCGAAAGCGCGCCCAACCGCAACCCGAAGCAG 1132
QY 1018 TTCTCCAGGCGCAGACGCTTTTGTATGTCACAGGAGCAGCGGCGTCCGCTACGCT 1077
DB 1133 TTCCGAGACCCCAACCGCTCCGACGTCAACCGGAGAACCCGCGCACTCTGTGTCGG 1192
QY 1078 AAGAGCCCCCATGCTGCCCCGGGGGTGTCCTTGTCCGCTCGAGCGAGATCGCCGTG 1137
DB 1193 CAGGCGATCCACTTCTGCAATGGGCGCGCTGGCCCAAGCTGAGGGGAGGTGGCGCTG 1252
QY 1138 GGACCATCTCTCGATGATTCGCCGAGATGAAGCTG 1173
DB 1253 CGGCGCTTTGGCGCGCTTCCCGCTGTGCGCTG 1288

RESULT 6
ABX56047

ID ABX56047 standard; DNA; 1347 BP.

AC ABX56047;

DT 13-FEB-2003 (first entry)

XX M. echinospora calicheamicin biosynthesis gene calW.

XX Calicheamicin biosynthetic gene cluster; aryltetrasaccharide;
KM aglycone; calicheamicin resistance; nonchromoprotein enediylne;
KM enediylne resistance; bone marrow cell; gene therapy; gene; ds.

XX Micromonospora echinospora spp. calichensis.

PN WO200279465-A2.

PD 10-OCT-2002.

PF 28-NOV-2001; 2001WO-US44285.

PR 28-NOV-2000; 2000US-0724797.

XX (SLOK) SLOAN KETTERING INST CANCER RES.

PI Thorson J;

DR WPI; 2003-092897/08.

DR P-PSDB; ABU11985.

XX Novel nucleic acid molecule from nonchromoprotein enediylne biosynthetic
PT gene cluster from Micromonospora echinospora useful for conferring
PT calicheamicin resistance on a subject -

XX Claim 9; Page 120-122; 179pp; English.

XX The present invention relates to the isolation of the Micromonospora
CC echinospora spp. calichensis calicheamicin biosynthetic gene cluster
CC encoding proteins and enzymes used in the biosynthetic production of
CC calicheamicin, including aryltetrasaccharide and aglycone. The gene
CC cluster also includes the gene encoding for the protein that confers
CC calicheamicin resistance. The calicheamicin biosynthetic gene cluster
CC is a nonchromoprotein enediylne biosynthetic gene cluster. Expression
CC vectors containing genes from the biosynthetic gene cluster are also
CC disclosed. The expression vectors are useful for producing calicheamicin
CC biosynthetic proteins. The calicheamicin self-resistance gene provides
CC an approach for gene therapy, for example, by introduction of enediylne
CC resistance genes into bone marrow cells, thus increasing resistance and
CC allowing tolerance to chemotherapeutic doses of calicheamicin.
CC ABX56028-ABX56073 represent genes from the M. echinospora calicheamicin
CC biosynthesis gene cluster.

SQ Sequence 1347 BP; 163 A; 505 C; 513 G; 166 T; 0 other;

Query Match

10.2%; Score 128.8; DB 25; Length 1347;

Best Local Similarity 49.7%; Pred. No. 4,2e-20;
Matches 356; Conservative 0; Mismatches 357; Indels 3; Gaps 1;

QY 459 CCCGATGCGCGCGATCAAGCGCTGTGTTGAAGTTCCGCGCAGTGTACAGAACTTCCG 518
DB 558 CTTGTGCGCGTGTCAAGTATCCGATCCGTCGCGGGTGAATCTCGAATCTCGCGCT 617
QY 519 TGGCTTGGCTCGGAGACTGCGCGCGCTCGCGCGTGGGTTGGTCCCGAGTGAATGA 578
DB 618 GCGTGAACGCGAAGCGGTTTCTTGACGCGCGGTTCACCGGATGTAGTGCAGGCT 677
QY 579 GAGACCAAGACCTGTGTGCGCTCCGTACAGCGAGGGGTGCGCTGTCCATGATGCTCT 638
DB 678 GCGCGCGAGACAGCGCGCGCGCGTGGCGCGAGATCCGCGAGTCTCTGAGCGGCTGT 737
QY 639 CGATAGCGCGCGAGAAACCGCGCTCGAAATGACGTCTTGACAGTCTCTTACAGCCGA 698
DB 738 GACCGCAAGAGAGCGGCAACC---CGGCGACGACGTGTCAAGCCGCTGTGCGCGCA 794
QY 699 GCGCGAGCGCGACGAGCTGAGCAGAGAGAGTGTGCGCGCTGCGGAGTATGCGC 758
DB 795 GCGCGCGCGCGCGAGCCCGACACAGAGCGCTGTGAGACATGCTCTGTCTGTCTGT 854
QY 759 TCGTGCACCGATACACAGATCTTATGCGCTGTGCTGCTCAACTGTCTGCGGTCT 818
DB 855 CCGCGCGACGTCACAGACGTGCAATGATCTGCTGAGCGTGTGCGCTGTGACCCA 914
QY 819 GCGCGAGCGCTCGAGCTGTGTGAAGCGCGAGCCGCGCTCATGAGAAACGCGCTGTGA 878
DB 915 TCGGAGCGGCTGCGCGCGCTGTGCGCGAGCGGAGCGGTTCCCGCGCTGTGAGGA 974
QY 879 GGTGCTCGCGCTTGCACATATCTCAGAAATGGAATGCGGTTTGGCAGGACAGACCT 938
DB 975 GCTGCTGCGGTACTTACCATGTGTGAGCGCGAGCGGAGACCGCGACCT 1034
QY 939 GAGTACTGCGCGGCGATCGATCAAGAAAGGAGATGCTTTCTCTGATCCGAGCGC 998
DB 1035 GACGCTCGGTGGGTTCACATCCGCGCGGAGGGGGTGTGGGCTGTGGCCAGGCGC 1094
QY 999 CCTGAGAGATGGAGCTGATTTCTCCAGGCGAGACGTGTTGATGTGCCAGGAGACAGG 1058
DB 1095 CAACCGGAGCCCGGCGGCGTGTGACCGGCGGACGATTCGACCGGAGACCGCGCGC 1154
QY 1059 CGCGAGCTCGCGGTACGTAGAGGCGCCCATGTGCGCGCGGAGTCCCTGTGCGCT 1118
DB 1155 GCACCACTCGCTTGTGCTACGAGACACATGTGCCCGGCGACACTGCGCGCT 1214
QY 1119 CGAGCGGAGATCGCGCTGTGGGACCACTTCCGTGAGTTCCCGAGATGAAGCTGA 1174
DB 1215 GGAAGTGAACGTGCGGCTGAGCGGCTGTGCGGCGGCTGCCGCGGTGA 1270

RESULT 7
AAA75635

ID AAA75635 standard; DNA; 5970 BP.

AC AAA75635;

DT 22-JAN-2001 (first entry)

XX Nucleotide sequence of ORF12 which encodes a transcriptional activator.

XX Naphenolide synthase; polyketide synthase gene; naphenolide polyketide;

KM antibiotic; C12-hydroxylase; pick; desosamine biosynthesis;

KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;

XX picromycin biosynthesis; ss.

OS Streptomyces venezuelae.

PN US6117659-A.

PD 12-SEP-2000.

XX

PF 27-MAY-1999; 99US-0320878.
 XX 28-MAY-1998; 98US-0087080.
 PR 22-SEP-1998; 98US-0100880.
 PR 08-FEB-1999; 98US-0119139.
 PR 20-MAY-1999; 99US-0134990.
 PR 30-APR-1997; 97US-0846247.
 PR 06-MAY-1998; 98US-0073538.
 PR 28-AUG-1998; 98US-0141908.
 XX
 XX (KOSA-) KOSAN BIOSCIENCES INC.
 PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
 DR WPI; 2000-610844/58.
 XX
 PT New recombinant pick hydroxylase gene of *Streptomyces venezuelae* useful
 PT for converting ketolides to antibiotics and as antibiotics and
 PT intermediates in the synthesis of compounds with pharmaceutical value
 PT
 PS Disclosure; Columns 41-44; 117pp; English.
 XX
 CC The present sequence is used to produce the recombinant DNA compounds
 CC of the invention. The specification describes a recombinant DNA compound
 CC expressing recombinant polyketide synthase genes in host cells for the
 CC production of nambonolide, nambonolide derivatives and polyketides that
 CC are useful as antibiotics and as intermediates in the synthesis of
 CC compounds with pharmaceutical value. The DNA compounds may also encode
 CC a C12-hydroxylase (pick), desosamine biosynthesis and desosaminyl
 CC transferase enzymes (useful for conversion of ketolides to antibiotics),
 CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
 CC These compounds are also useful for increasing the antibiotic activity
 CC of a compound relative to the unhydroxylated compound. The recombinant
 CC host cells are useful as genetic systems that allow rapid engineering
 CC of the nambonolide polyketide synthase. These would be valuable for
 CC creating novel ketolide analogs for pharmaceutical applications.
 CC
 SQ Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;

Query Match 9.8%; Score 123; DB 21; Length 5970;
 Best Local Similarity 47.5%; Pred. No. 1.2e-18;
 Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;

QY 236 AGTACTCGGCGCAATCCCGAGCTCAGGATATGAAGAATGTTGCGGCTGC 295
 DB 1582 ACTCAAGACTCCCTGACCGGAGCCGCGCTCAACCAACATGCTGGAGTCCG 1641
 QY 236 CGCCGAGAGATCACCTCGGCTCCGCAAGCTGTCACCCCTGTTTACGTACCGCGCA 355
 DB 1642 ACCGCGCGGCGACACCGGCTGCGCAAGCTGTGGCCCTGTGATTCACATGCCCGGG 1701
 QY 356 TCGACCTGCTGCGCGCCGAAATACAGCGCACCTGACCAAGTGTCTGCTGCTCCG 415
 DB 1702 TCGATTTCTGCGGCGCGGCTCCAGAGATGTCAGCGGCTCGTGAAGCCATGTGG 1761
 QY 416 GACAAGAGAGATTGACGTTGTCGGGATTAAGCGGAGGATCCCGATGCCCGGATCA 475
 DB 1762 CGGCGCGCGAGCGCGCGCGGATGATGATGATCTCTGCGCTGGC---GATATCA 1818
 QY 476 GCGCTCTGTGAAGGTTCCGCGCGGATGACGAGAAATTCGTGCTTCCGCTCGGCGA 535
 DB 1819 CCGTATCTCCGAATCTCTCGGCTGCGCGAGCGGAGCCGCGCTTCCGCTGTGGA 1878
 QY 536 CTGCGCGCGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595
 DB 1879 CCGAGCGCTTCTGCTTCCGCGAGATCCCGCCAGGCCAGACCG----- 1923
 QY 596 TCGGCTCGGATACGAGGGGCTCGGCTGCTCATGAGAGTCTCATGAGAGGGCGAGGA 655
 DB 1924 -----CATGCGCGAGATGAGCGGCTATCTTCCGGCTCATTCAGATCCAGCGGGC 1977
 QY 656 ACCCGCTGAAATGACGTCTTGAAGATGCTTCAAGCCAGGCGGAGCGGAGAGGCG 715

DB 1978 AGAAGCGGAGAGACTGCTCAGCGGCTGTCGCGGATCCAGGAGAGAGACGCTCCCGC 2037
 QY 716 TGAGACGAAGAGCTGCTGCGGCTCGGAGGATGATGCTGTGTCGACCCGATACCA 775
 DB 2038 TGACCTCCGAGAGTGTCTCGATGAGCCATCTGCTGTGTCGCGGAGAGAGACCA 2097
 QY 776 CGATTCACCTTATGCGCTGCTGCTGCTCAACCTGCTGCTGCTGCTGCTGCTGCTGCT 835
 DB 2098 CGGTCAATCTATGCGGATGAGGAGATGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 2157
 QY 836 TGTGAGAGCGGAGCGCGGCTCATGAGAAAGCGGCTCATGAGAGTGTCTGCTTGCACA 895
 DB 2158 CCTGCGGCGGAGATGACGCTCTTGAAGCGGCTGAGAGAGATGTTGCGCTACGAGG 2217
 QY 896 ATATCTCAGATAGGAATGGAATGCTGCTTCCGAGGAGAGACTGAGATGCTGCGGCGAT 955
 DB 2218 GCGGCTGGAATCCGCGGCTTACCGCTTCCGCTGAGCCGCTGAGAGAGAGAGAGAG 2277
 QY 956 CGATCAAGAAAGGAGAGATGCTTCTCTCTGATCCGAGGCGCTGAGAGATGAGAGACTG 1015
 DB 2278 TCATCTCCGCGGCTGACAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2337
 QY 1016 TATCTCCAGGCGAGAGCTGTTGATGTGACAGGAGACGAGGCGGAGCTTCCGCTGACG 1075
 DB 2338 GCTTCCGAGACCCGACCGCTTTCAGATCCGCGGAGACCGGCGGAGCTTCCGCTTCCG 2397
 QY 1076 GTAGAGGCGGCGGAGTGTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1135
 DB 2398 GCCAGGATTCATCTTCTGATGAGGCGGCTTGTGCGGCTTGTGAGGCGCGGAGTCCG 2457
 QY 1136 TGGGACCATCTTCCGATGATGCTCCCGAGATGAGCTGAAGAAATCTCCG 1186
 DB 2458 TCCGCGCTTCTGGAAGCTGCGGAGCTTCCGCTGAGAGTCTCTCCCG 2508

RESULT 8
 AA256003
 ID AA256003 standard; DNA; 5970 BP.
 XX
 AC AA256003;
 XX
 DT 23-MAR-2000 (first entry)
 XX
 DE Contig 002 from cosmid pXOS023-27 from *Streptomyces venezuelae*.
 XX
 KW Nambonolide polyketide synthase; PKS; cosmid pXOS023-27; contig 002;
 KW ketolide; amino transferase dehydratase; hydroxylase; picromycin;
 KM antibiotic production; nambomycin; da.
 OS *Streptomyces venezuelae*.
 XX
 FH Key
 FT CDS location/Qualifiers
 FT CDS complement (1..995)
 FT FT /*tag= a
 FT FT /product= PICIV
 FT FT /note= "Partial amino transferase-dehydratase"
 FT FT /transl_except= (Pos:180..182, aa:Xaa)
 FT FT 1356..2606
 FT CDS /*tag= c
 FT FT /product= c
 FT FT /note= "Cytochrome P450 hydroxylase"
 FT FT 2739..5525
 FT FT /*tag= d
 FT FT /product= [transcriptional activator
 FT FT /transl_except= (Pos:4818..4820, aa:Xaa)
 FT FT /note= "Xaa = unknown"
 PN MO9961599-A2.
 XX
 PD 02-DEC-1999.
 XX
 PF 27-MAY-1999; 99WO-US11814.

```
XX 28-MAY-1998: 98US-0087080.
PR 28-AUG-1998: 98US-0141908.
PR 22-SEP-1998: 98US-0100880.
PR 08-FEB-1999: 99US-0119139.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
PI Ashley G, Belach MC, Belach M, McDaniel R, Tang L;
DR WPI; 2000-072618/06.
DR P-PSDB; AAY67209, AAY67213, AAY67218.
XX
PT New recombinant DNA encoding a domain of narbonneolide polyketide
PI synthase, for production of ketolide antibiotics.
XX
PS Disclosure; Page 35-37; 98pp; English.
XX
CC This is contig 002 from the recombinant cosmid PKOS023-27 DNA sequence
CC (see AA556001) which contains a Streptomyces venezuelae DNA insert. The
CC cosmid contains open reading frames which encode the various modules of
CC the narbonneolide polyketide synthase (PKS). The invention relates to
CC recombinant DNA containing a coding sequence for a narbonneolide PKS.
CC Polyketides are compounds synthesized from 2-carbon units through a
CC series of condensations and subsequent modifications. Modular PKSs are
CC responsible for the production of many antibiotics including picromycin.
CC The narbonneolide PKS consists of a loading module, six extender modules,
CC and two thioester domains. Four proteins make up the narbonneolide PKS
CC (PICAI, PICAI, PICAI, and PICAI). PICAI includes the loading module
CC and extender modules 1 and 2, PICAI includes extender modules 3 and 4,
CC PICAI includes extender module 5 and PICAI includes extender module 6
CC and a type II thioesterase domain. The second type II thioesterase
CC domain is found on the PICB protein. The nucleotide sequences encoding
CC all of these proteins can be isolated in recombinant form from the
CC recombinant cosmid PKOS023-27. Narbonneolide is desosaminylated in S.
CC venezuelae to yield narbonneolide, the desosaminyl transferase enzyme is
CC required for this conversion, and the desosaminyl biosynthetic genes are
CC also found in cosmid PKOS023-27. The recombinant DNA of the invention is
CC used to express, in transformed cells, narbonneolide (or its derivatives)
CC or other ketolides (particularly hybrids), which may then be converted.
CC (e.g. by other enzymes recombinantly expressed in the same hosts) to
CC polyketide antibiotics or their intermediates. The antibiotics are useful
CC in human or veterinary medicine.
XX
SQ Sequence 5970 BP, 806 A, 2219 C, 2142 G, 800 T, 3 other;
Query Match 9.8%; Score 123; DB 21; Length 5970;
Best Local Similarity 47.5%; Pred. No. 1,2e-18;
Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;
QY 236 AGTACTGTCGCGCATTCCTCCAGCTCAGGATATGAGAGTACGATTTCTCGGCTGC 295
DB 1582 ACTCCACGACTCTCCCTGACCGGCGCGCGCTCAACACACATGCTGAGTCCG 1641
QY 296 CGCGGAGATCAGCTGCGGATCCGAGCTGTCAACCGCTGTATGATGACGCGCA 355
DB 1642 ACCCGCGGCGACATCCGCTGCGAGCTGTGCGCTGTGATTCACATGCGCGG 1701
QY 356 TCGACTGTCGCGCGCGGAAATACAGCGACCGTGCACGAGCTGTGCTGCTCGC 415
DB 1702 TCGAGTGTCTGCGCGCGCGGATCGAGATGCTGAGCGGCTGTGCGCGCTGC 1761
QY 416 GACACAGAGAGTTCAGCTGTGCGGATTAAGCGAGGAAATCCGATGCGCGCATCA 475
DB 1762 CGGCGCGCGCGCGCGCGCGATGATGAGATCTCTGCGCTGCGCGCTGC---CGATCA 1818
QY 476 GCGCTGTGTAAGTTCGCGCGAGTGTGAGAGAGAGTTCGCTGCTGCTGCGCGCA 535
DB 1819 CCGTATCTCCGACTCTCTGCGGCTGCGCGAGCGCGCGCTTCCGCTTGTGA 1878
QY 536 CTGCGCGCGCGCTGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595
DB 1879 CCGAGCGCTTCTGCTTCTGCGGAGATCCGCGCGCGCGCGCGCGCGCGCGCGCG 1923
```

```
QY 596 TCGCTGCTGACCGAGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 655
DB 1924 -----CCATGCGCGAGATGAGCGCGCTATCTCTCCGCGCTATCTCAACCGCGCGC 1977
QY 656 ACCCGCTGAAATACGCTTGTGAGAGTGTGCTTCAAGCGCGCGCGCGCGCGCGCG 715
DB 1978 AGGAGCGGAGAGAGCTGCTGAGCGCGCTGCTGCGGACAGAGAGAGAGAGAGAG 2037
QY 716 TGAGCAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 775
DB 2038 TGACCTCGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2097
QY 776 CGATCTACTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 835
DB 2098 CGGTAACTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2157
QY 836 TGTGAGAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 895
DB 2158 CCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2217
QY 896 ATATCTGAGATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 955
DB 2218 GCGCGGTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2277
QY 956 CGATCAAGAGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1015
DB 2278 TCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2337
QY 1016 TATCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
DB 2338 GCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2397
QY 1076 GTAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1135
DB 2398 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2457
QY 1136 TGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
DB 2458 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2508
RESULT 9
ID ABS56092 standard; DNA; 5970 BP.
XX
XX ABS56092;
AC 21-JAN-2003 (first entry)
XX
XX
XX Contig 002 from cosmid PKOS023-275 containing S. venezuelae DNA insert.
DE Narbonneolide polyketide synthase; PKS; desosaminyl biosynthetic gene;
XX desosaminyl transferase gene; beta-glucosidase gene; antibiotic;
XX pick hydroxylase gene; C12 hydroxylase gene; narbonneolide;
XX desosaminylated polyketide; narbonneolide biosynthesis; mutant;
XX picromycin biosynthesis; de.
XX
XX Streptomyces venezuelae.
OS Synthetic.
OS
OS WO200297062-A2.
PN
XX
XX 05-DEC-2002.
PD
XX
XX 22-FEB-2002; 2002WO-US05642.
PF
XX
XX 22-FEB-2001; 2001US-0793708.
PR
XX
XX (KOSA-) KOSAN BIOSCIENCES INC.
PA
XX Ashley G, Belach MC, Belach M, McDaniel R, Tang L;
XX
```


DR WPI: 2003-041412/03.
 XX Preparation of polyketides by recombinant DNA technology, useful as
 PT antibiotics and as intermediates in the synthesis of pharmaceutical
 PT compounds -
 XX
 PS Disclosure: Page 40-41; 127p; English.
 XX
 CC The present invention relates to recombinant DNA sequences encoding
 CC for a naphtholide polyketide synthase (PKS) domain, and methods of
 CC producing polyketides by recombinant DNA technology. The recombinant
 CC DNA sequences are derived from Streptomyces venezuelae desosamine
 CC biosynthetic, desosaminyl transferase, beta-glucosidase, or pick
 CC (C12) hydroxylase genes. The method is useful for transforming a cell
 CC with a recombinant expression vector that encodes a functional
 CC beta-glucosidase gene, and therefore for increasing the yield of a
 CC desosaminylated polyketide in a cell. The recombinant methods and
 CC materials are useful for expressing polyketides with significant
 CC antibiotic activity, derived in whole or in part from the naphtholide
 CC PKS gene, and other genes involved in naphthomycin and picrocin
 CC biosynthesis in recombinant host cells. The present sequence
 CC represents contig 002 from cosmid pKOS023-2/5 containing
 CC S. venezuelae DNA.
 CC
 XX Sequence 5970 BP; 806 A; 2219 C; 2142 G; 800 T; 3 other;
 XX
 Query Match 9.8%; Score 123; DB 25; Length 5970;
 Best Local Similarity 47.5%; Pred. No. 1.2e-18;
 Matches 452; Conservative 0; Mismatches 475; Indels 24; Gaps 2;
 QY 236 AGTACTGTCGCGCATTCCTCCAGCTCAGCATATGAAAGATGCGATTGTCGGGCTGC 295
 DB 1582 ACTCCAGACATCCCTGACCCGAGGCGGCGGCTCAACCAACATGCTGAGTCCG 1641
 QY 296 CGCGGAGATCAGCTGCTGGGTCGCAAGCTGTAACCCGCTTTTACGTCAGCGCCCA 355
 DB 1642 ACCCGCCCGGACACCCGCGCTGCGCAAGCTGCGCCCGCTGATGACATGCGCCGG 1701
 QY 356 TCGACCTGTCGGCGCGCAAAATACAGGCGCATGTCGATGCTGATGCTGCTCG 415
 DB 1702 TCGAGTTGTCGGCGCCCGGCTCCAGAGATGTCGAGGCTGCTGAGACGCTGCTGG 1761
 QY 416 GACAAGAGAGATGTCAGCTTGTGCGGATTAAGCGAGAGATCCCGATGCGCGGATCA 475
 DB 1762 CGGGGCCCGAGCGCGCGCCGATGATGAGATCCCTGCGCGCTGC---CGATCA 1818
 QY 476 GCGCTGTGTAAGTTCGCGCCGATGTGACGAGAATTCGTCGCTTCGCTCGCGCA 535
 DB 1819 CCGTATCTCCGAATCTCTCGGCGTCCGAGCCGAGCCGCGCGCTTCGCGCTGGA 1878
 QY 536 CTGGCGCGCGCTGCGCGTGGGTTTGGTCCCGAGGTGATGAGAGACAAAGACCTGG 595
 DB 1879 CCGAGCGCTTGGCTCTCCGAGCATCCCGCCCAAGCCG----- 1923
 QY 596 TCGCGTCCGTCACGAGGCGCTCGGCTGTCATGAGCTCTTCATGAGGCGCGCAGA 655
 DB 1924 -----CCATGGCGAGATGAGCGGCTATCTCTCCGCTATATGATCCCAAGCGCGG 1977
 QY 656 ACCCGCTGAAAATGACGCTTTGACGATGCTGCTTCAAGCGGAGCCGACGCGACAGG 715
 DB 1978 AGGACGCGAGAGACCTGCTCAGCGCGCTGTCGAGACAGAGACGAGGCTCCCGG 2037
 QY 716 TGAGACGAGAGAGTGTGCGCTGCGCTGCGGTGATGATGCTGCTGCGACCATTA 775
 DB 2038 TGAACCTCGAGAGTGTGCTGATGAGCCACATCTCTGTCGCGGCGGACGAGACCA 2097
 QY 776 CGATCTACTTATCGCTTCCGCTGTCTCAACTTCGCTCGGTCGCGGAGGCTCGAGC 835
 DB 2098 CGGTAAATCTATGCGCAACGCGCATGTAGCGCTCTCTCGACCCGACAGCTGCGG 2157
 QY 836 TGTGTAAAGCGGCGCGGCTCATGAGAGAACGCGCTGATGAGTGTCTCGCTTCGACA 895
 DB 2158 CCTGCGGCGGACATGACGCTCTTGTGACGCGCGGCTGAGAGATGTTGGCTTCGAGG 2217

QY 896 ATATCTCAGAAATAGAACTGTGCTTTCGCCAGGACCTGAGTACTGCGGGCAT 955
 DB 2218 GCCGGGTGAATCGCGACCTACGCTTCCCGGTGAGCCCGTGAACCTGAGCGGACGG 2277
 QY 956 CGATCAAGAAAGGAGATGCTTTTCTCTGATCCGAGCGCTCTGAGATGAGACTG 1015
 DB 2278 TCATCCCGCGCGGAGACAGGCTCTGCTGCTGCGGACGACCGACCCCGGAGC 2337
 QY 1016 TATCTCCAGGCGCAGAGTGTGATGAGAGGAGACAGGGGCGGAGCCTCGGTACG 1075
 DB 2338 GCTTCCGAGCCGACCGCTTGCACATCGCGCGGACACGCGCCCATCTCGCTTCG 2397
 QY 1076 GTAGACCCCGCATGTGCGCGCGGGTGTCTTGTGCTGCGCTCGAGCGAGATCGCG 1135
 DB 2398 GCCAGGATCATCTTGTGATCGGCGCCCGCTTGCGCGGTTGAGGCGCGGATCGCG 2457
 QY 1136 TGAGCAACCATCTTCCGATGAGTTCGCCAGATGAAGTGAAGAAATCTCCG 1186
 DB 2458 TCCGCGCCCTTTCGAACGCTGCCGAGCTGCGCTGACGCTCTCCCG 2508

RESULT 10
 ABN8910
 ID ABN8910 standard; DNA; 1191 BP.
 XX
 AC ABN8910;
 XX
 DT 21-AUG-2002 (first entry)
 XX
 XX Streptomyces sp. cytochrome P450 encoding DNA SEQ ID NO.1.
 DE Streptomyces sp. cytochrome P450, anticancer; UCN-01, monooxygenase;
 KW Streptomyces; cytochrome P450, anticancer; UCN-01, monooxygenase;
 KW microorganism; gene; ds.
 XX
 OS Streptomyces sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1191
 FT /tag= a
 FT /product= "cytochrome P450"
 FT
 XX JP2002058490-A.
 XX
 XX 26-FEB-2002.
 XX
 XX 22-AUG-2000; 2000JP-0251304.
 XX
 XX 22-AUG-2000; 2000JP-0251304.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX WPI: 2002-410015/44.
 DR P-PSDB; ABB81315.
 PT Novel cytochrome P450 used for improving the productivity of anticancer
 PT drug UCN-01 -
 XX
 PS Claim 5; Page 9-11; 22pp; Japanese.
 XX
 XX The present invention describes 3 Streptomyces cytochrome P450 (1)
 CC sequences, ABB81315 to ABB81317 encoded by ABN8910 to ABN8912
 CC respectively. (1) have monooxygenase activity and improve the
 CC productivity of an anticancer drug designated UCN-01. (1) is used
 CC for the production of UCN-01 with transformed Streptomyces sp.
 CC microorganisms.
 XX
 SQ Sequence 1191 BP; 169 A; 423 C; 431 G; 168 T; 0 other;
 Query Match 9.7%; Score 122.6; DB 24; Length 1191;
 Best Local Similarity 48.3%; Pred. No. 1.1e-18;
 Matches 430; Conservative 0; Mismatches 434; Indels 27; Gaps 2;

QY 298 CCGGAGGATCAGCTCGGGTCCGGAAGTCGTCACCCGGTCGTTTACGTCAGCCGCGATC 357
 DB 271 CCGCCCGACACACACCCGGCTGGCGGGCTGATGCGCCGGAGTTCACCGGACGCGGGGTC 330
 QY 358 GACCTGCTGGCGCGGCAAAATAGACGCAACGTCGACAGCTGCTGATGCTGCTCCGA 417
 DB 331 GAGCGGCTGGCGCGGATCCAGAGATCACGAGTCTGCTGAGACGAGATGCTGCGG 390
 QY 418 CAAAGAGATTGACGCTTGTGCGGATTAACGCGAGGAGAAATCCGATGCGCGGATCAGC 477
 DB 391 CCGCGCGCGCGGACCTGCTGAGTCTTCTGAGTACCCGCTGCTGACCGTCACTCTGC 450
 QY 478 GCTGTTGAAGTTCCGCGCGAGTGTAGAGAAATTCGCTGCTGCGGCTGCGGAGT 537
 DB 451 GAATCTCTGGGGTCTCCGAGA-----TCGACCGGGGCGCC 486
 QY 538 GCGCGCGCGCTGGCGCTGGATTGTGATCCCAAGTTCAGTAGAGAGACAAAGCCCTGTATC 597
 DB 487 TTCGCGAAGCTGTCAGACGAGAGGGGTGGACCCACAGGCGGGAGAGCGAGTAC---GGC 543
 QY 598 GCGTCCGTCAACGAGGGGCTCGCGCTGCTTCATGACGTCCTCGATGACGCGCGAGAAC 657
 DB 544 GCTTCGTCAACTCGCCGCTTACCTGAGAGATTGTGAGAGAGAAAGCGTGCACCGC 603
 QY 658 CCGCTCGAAATGAGCTTGTGACGATGCTTTCAGCGCGGCGGACGCGCGAGCGGCTG 717
 DB 604 CCGGCGGAGATCTGCTGACCGGCTGATTCGGAACAGACGAGGAGCGGACGCTGTCG 663
 QY 718 AGCAGCAGAGAGTGTGCTCGCTCGTGGGTGCGATTATGCTGCTGCGACCGATACAG 777
 DB 664 TCACCCGGGAGTGTGCGGCGCATGCGCTTCATCTGCTGATGCGCGGACGAGACAC 723
 QY 778 ATCTACCTTATGCGCTTCCGTGCTCAACCTGCTCGGTCGCGCGGCGCTCCAGCTG 837
 DB 724 GTCAACCTCATCACCGGTGCGCTGCGCTTCCTCACCATCTCCGCGCAACTCGCCAG 783
 QY 838 GTGAAGCGCGAGCGCGGCTCATGAGAAAGCGCTGATGAGGTGCTCGCTTCGACAT 897
 DB 784 GTGCGGGGCGACATGAGCTTCTGACGCGGCTGTGAGAGAGACGCTGCGCCATGAGGG 843
 QY 898 ATCTCAGAAATGAACTGTGCGTTCGCGCAGGACAGACCTGAGATTAATCGCGGCGATC 957
 DB 844 CCGGTGAGAAAGCGCGATTCGCGCTTCGCGCGGACCGCTGAGATGAGGGGCGACGTC 903
 QY 958 ATCAAGAAAGGAGATGCTTCTTCTGATCCGAGCGCTTGAAGATGGAATGTA 1017
 DB 904 ATCCAGCGCGGAGACCGGTGCTGATCGGCGTGGCGCGCGCATGCGGACCGCGCGC 963
 QY 1018 TTCTCCAGGCGCAGACGTTTGTATGTGCGACGCGGACAGCGGCGCGCTCGGCTACG 1077
 DB 964 TATCCCGGCGCGACCGCTTCACATTCACCGGAGACACGCGGCGATCTCGCTTCGCG 1023
 QY 1078 AGAGCCCGCATGTCTGCGCGGCGGTGCTTCTGCTCGCTCGAGCGGAGATCGCGCTG 1137
 DB 1024 CACGGGATCATCTTCTGCGCGCGCGCGCTGCGCGCTGAGGCGCGGGGTGCTG 1083
 QY 1138 GGCACCATTTCCGTGATGTTCCCGAGATGAAAGCTAAAGAACTCCGCTG 1188
 DB 1084 CCGGCGCTCTGAGAGCGGTGCTGCTACGCGCGGACGCGGCGACCGGCG 1134

RESULT 11

AAZ87301 standard; DNA; 1251 BP.

AAZ87301;

05-JUN-2000 (first entry)

S. venezuelae macrolide biosynthetic gene plkC, SEQ ID NO:38.

DE Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 42

KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; ashtma;
 KM chronic obstructive pulmonary disease; respiratory inflammation;
 KM hypercholesterolemia; crop protection agent; ds.
 OS Streptomyces venezuelae ATCC15439.
 PH Key Location/Qualifiers
 FT CDS 1..1251
 FT /tag= a
 FT /product= "plkC"
 FT /transl_except= (pos:307..309, aa:leu)
 PN W020000620-A2.
 PD 06-JAN-2000.
 PF 25-JUN-1999; 99WO-US14398.
 PR 26-JUN-1998; 98US-0105537.
 PA (MIND) UNIV MINNESOTA.
 PI Sherman DH, Liu H, Xue Y, Zhao L;
 DR MPI: 2000-160679/14.
 DR P-PDB: AAY77196.
 PT Desosamine and macrolide biosynthetic gene clusters, useful for, e. g.
 PT synthesis of methymycin and pikromycin -
 PS Claim 15; Page 428-429; 438pp; English.
 XX The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the eryC gene cluster of *Saccharopolyspora erythraea* or
 CC streptomycetes antibiotics. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics.
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants.
 CC Sequences AAZ87295-287302 represent macrolide biosynthetic genes from
 CC Streptomyces venezuelae ATCC 15439, which encode proteins
 CC AAY77190-77197.
 XX
 XX
 SQ Sequence 1251 BP; 171 A; 484 C; 413 G; 183 T; 0 other;
 Query Match 9.6%; Score 121.4; DB 21; Length 1251;
 Best Local Similarity 50.4%; Pred. No. 2.1e-18;
 Matches 295; Conservative 0; Mismatches 290; Indels 0; Gaps 0;
 QY 602 CCGTACCGAGGGGCTCGCGCTGCTCCATGAGCTGCTGATGAGGCGGCGAGAACCGGC 661
 DB 569 CCATGCGCGAGATGAGCGGCTATCTCTCCGCGCTATGATCCAAAGCGGCGGAGAGC 628
 QY 662 TCGAAATGACCTCTTGAAGATGCTGCTTCAGCGCGGAGCGGACGCGACGAGCTGAGCA 721

```

Db 629 GCGAGGACCTGCTACGCGCGCTGTCGCGACCGAGCAGACGCGCTCCGCGCTGACCT 688
Qy 722 CGAAGAGCTGCTCGCGCTCGTGGGTGCGATATCGCTGCTGCGACCGATACGATCT 781
Db 689 CGAGGAGCTGCTCGCTGATGCGCCACATCTCTGCTGCGGGGACGACGACGATCTCA 748
Qy 782 ACTTATGCGCTGCTGCTGCTCAACCTGCTGCGGCTCGCGCGCTGAGCTGATGA 841
Db 749 ATCTGATCGCCAAACGCGATGTCGCGCTGCTGCTGACACCCGACAGCTGGCCGCTGCG 808
Qy 842 AGCGCGACCGCGGCTCATAGAGAACCGGCTCGATGAGTGTCCGCTTCCGACATATCC 901
Db 809 GGGCGGACATGACGCTCTTGAACGCGCGGCTGAGAGATGTTGGCTTACGAGGCGCCGG 868
Qy 902 TCAGATAGAACTGTGCTTTCGCGAGGACGAGCTGAGATGCGGGGCGATGATCA 961
Db 869 TGGAAATCGCGACCTACCGCTTCCGGTTCGAGCCCTGACCTGAGACGCGACGCTATCC 928
Qy 962 AGAAGGGGAGATGCTTTCTCTGATCCGAGCGCCCTGAGAGATGGGACTGATTTCT 1021
Db 929 CGCGCGGTGACACGGTCTCTGCTGCTGCTGCGCGACCGACCGCCCGAGCGCTTCC 988
Qy 1022 CAGAGCCAGACGTGTTTATGTCGACGCGGACCGCGGCGGAGCTTCCGCTACGTAAG 1081
Db 989 CGAACCACCGACCTTGCACATCCGCGGAGACCGCGCCATCTCGCTTCCGCGCACG 1048
Qy 1082 GCGCCCATGCTGCGCGCGGGGTGCTTGTGCTGCGCTCGAGGCGGAGATCGCGCTGAGCA 1141
Db 1049 GCATCCACTTCTGATCGGCGCGCCCTTGGCCCGGTTGAGGCGCGGATCGCGCTCCGG 1108
Qy 1142 CCATCTTCGTAAGTTCCTCCCGAGATGAAGCTGAAGAAACTCCCG 1186
Db 1109 CCTCTTCGAACGCTGCGCGGACCTCGCGCTGGAAGTCTCCCG 1153

```

RESULT 12

ABZ7672
ID ABZ7672 standard; DNA; 1992 BP.

AC ABZ76672;

DT 30-APR-2003 (first entry)

DE Streptomyces sp. TM-7 related compactin DNA SEQ ID NO.1.

XX Streptomyces sp. TM-7; pravastratin; compactin; hyperlipidaemia;

KW antilipaeamic; microorganism; gene; ds.

OS Streptomyces sp. TM-7.

XX Key Location/Qualifiers

FT CDS 544..11761

FT /*tag= a

FT /product= "compactin related protein (ABP95987)"

FT 1782..1973

FT /*tag= b

FT /product= "compactin related protein (ABP95988)"

XX MO20029109-A1.

XX 12-DEC-2002.

XX 30-MAY-2002; 2002MO-JP05252.

XX 01-JUN-2001; 2001JP-0166412.

XX (SAOC) MERCIAN CORP.

XX Fujii T, Hirose S, Aritoku Y, Morimiyu T, Jondo O, Isehiki K;

XX WPI; 2003-148672/14.

XX P-PSDB; ABP95987, ABP95988.

PT Novel Streptomyces sp. produced polypeptide for hydroxylation of
PT compactin at 6beta-position and its encoded DNA, applicable in
PT constructing transformant microbes to synthesize pravastratin for
PT treating hyperlipidaemia -

PS Claim 1; Page 45-48; 67pp; Japanese.

XX The present sequence represents a DNA sequence which contains a base
XX sequence from bases 544-1758 in the sequence of (1) with 1992 base pairs,
CC or a DNA hybridisable with the DNA under stringent conditions and
CC encoding a polypeptide with hydroxylase activity on compactin at
CC 6beta-position. Also described: (1) DNA containing base sequences from
CC bases 544-1758 and from bases 1782-1970 in the sequence of (1) or a DNA
CC hybridisable with the DNA under stringent conditions and encoding a
CC polypeptide with hydroxylase activity on compactin at the 6beta-position;
CC (2) a polypeptide encoded by any of the DNA or containing an amino acid
CC sequence based on the polypeptide but with some amino acids deleted,
CC substituted or added and having hydroxylase activity on compactin at
CC the 6beta-position; (3) a recombinant DNA obtained by integrating with
CC any of the DNA; (4) a microorganism transferred with the recombinant DNA;
CC (5) a process for producing pravastratin by culturing the transformant
CC microorganism before isolating the culture liquor or cells, and addition
CC of compactin for reaction to give pravastratin for recovery; and (6)
CC Streptomyces sp. TM-6 (FERM BP-8002) or TM-7 (FERM BP-8003). (1) has
CC antilipaeamic activity. The polypeptide and its encoded DNA are applicable
CC in constructing transformant microorganisms to synthesize pravastratin for
CC treating hyperlipidaemia. With the recombinant microorganisms,
CC pravastratin can be produced efficiently, with much less galpha
CC hydroxylated epimer formed.

XX Sequence 1992 BP; 275 A; 691 C; 726 G; 300 T; 0 other;

XX Query Match 9.6%; Score 120.8; DB 25; Length 1992;

XX Best Local Similarity 50.7%; Pred. No. 3.1e-18;

XX Matches 290; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

```

Qy 602 CCGTACCGAGGAGCTCGCGCTGCTCCATGACGTCTCTGATGAGCGCGCAGAACCCGC 661
Db 1112 CCGCCCGGAGAGACTGAGAGAGCTATCTGGCGCTGCTGATGACGCTGAGCGAGT 1171
Qy 662 TCGAATGACGCTTTCGATGCTGCTTTCAGGCGGAGCGGACGAGGCTGAGCA 721
Db 1172 CCGCGCGGCGCTCTGTCGACGCTGTCAGAGGACGCTGAGAGGCGCGATCGAC 1231
Qy 722 CGAAGAGCTGCTGCGCGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781
Db 1232 GCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1291
Qy 782 ACTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
Db 1292 CGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1351
Qy 842 AGCGCGAGCGCGGCTCATGAGAACGCGCTCGATGAGTGTCTCGCTTCAATATCC 901
Db 1352 GCGCGGATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1411
Qy 902 TCAGATAGAACTGTGCTTTCGCGAGGACGCTGAGATGCTGCGGGGCGATGATCA 961
Db 1412 CCGACATGCGCGCGGCGGATGCGAGCGGACATGAGATGACGCGGACGCGCATCC 1471
Qy 962 AGAAGGGAGATGCTTTCCTGATCCGAGCGCGCTGAGAGATGAGATGCTGATTTCT 1021
Db 1472 GGGGGGGGAGGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1531
Qy 1022 CAGAGCCAGACGTGTTGATGTCGACGCGGACGCGGCGGAGCTTCCGCTACGTAAG 1081
Db 1532 CCGACCGGAGCGCTTGCACGCTGCGGCGGCGGCGGCGGACCACTGCGCTTCCGCTACG 1591
Qy 1082 GCGCCCATGCTGCGCGGGGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141
Db 1592 GGGTGCATGAGTCTGCGCGGAGAACCTGCGCGCTGGAAGTGAAGTCTACCGG 1651
Qy 1142 CCATCTTCGTAAGTTCCTCCGAGATGAAGCTG 1173

```

```

Db      1652 CCCTGTTGAGCGGCTGCCGCTCTCGGCTG 1663
RESULT 13
ID      ABV75558/c
XX      ABV75558 standard; DNA; 36538 BP.
AC      ABV75558;
XX      22-JAN-2003 (first entry)
XX      Saccharopolyspora butenyi-spinosyn biosynthetic gene cluster 2.
DE      Saccharopolyspora butenyi-spinosyn biosynthetic gene cluster 2.
XX      Butenyl; biosynthetic enzyme; PKS; polyketide synthetase; macroliide;
KW      metabolite; spinosyn; gene; ds.
XX      Saccharopolyspora sp.
OS      Saccharopolyspora sp.
XX      Key
FH      Location/Qualifiers
FT      complement (114..938)
FT      /tag= a
FT      /product= "bush"
FT      /note= "No start codon given"
FT      1389..2561
FT      /tag= b
FT      /product= "bush"
FT      2601..3353
FT      /tag= c
FT      /product= "bush"
FT      complement (3359..4546)
FT      /tag= d
FT      /product= "bush"
FT      4684..6303
FT      /tag= e
FT      /product= "bush"
FT      6317..7510
FT      /tag= f
FT      /product= "bush"
FT      7555..8406
FT      /tag= g
FT      /product= "bush"
FT      /note= "No start codon given"
FT      8640..9572
FT      /tag= h
FT      /product= "bush"
FT      /note= "No start codon given"
FT      complement (9668..10666)
FT      /tag= i
FT      /product= "bush"
FT      complement (10675..12135)
FT      /tag= j
FT      /product= "bush"
FT      complement (12864..14177)
FT      /tag= k
FT      /product= "bush"
FT      14627..15970
FT      /tag= l
FT      /product= "bush"
FT      /note= "No start codon given"
FT      16008..17144
FT      /tag= m
FT      /product= "bush"
FT      /note= "No start codon given"
FT      17168..17917
FT      /tag= n
FT      /product= "bush"
FT      complement (18520..19932)
FT      /tag= o
FT      /product= "ORF LI"
FT      complement (19978..20488)
FT      /tag= p
FT      /product= "ORF LI"

```

```

FT      /note= "No start codon given"
FT      complement (20536..21033)
FT      /tag= q
FT      /product= "ORF LIII"
FT      /note= "No start codon given"
FT      21179..21925
FT      /tag= r
FT      /product= "ORF LIV"
FT      complement (22671..23453)
FT      /tag= s
FT      /product= "ORF LVI"
FT      complement (23687..24886)
FT      /tag= t
FT      /product= "ORF LVII"
FT      complement (26177..26923)
FT      /tag= u
FT      /product= "ORF LVIII"
FT      /note= "No start codon given"
FT      27646..28476
FT      /tag= v
FT      /product= "ORF LIX"
FT      /note= "No start codon given"
FT      MO200279477-A2.
FT      10-OCT-2002.
FT      28-MAR-2002; 2002MO-US09968.
FT      30-MAR-2001; 2001US-280175P.
FT      (DOWC ) DOW AGROSCIENCES LLC.
FT      Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
FT      Mitchell JC;
FT      MPI: 2003-058434/05.
FT      P-REDB; ABP57683, ABP57684, ABP57685, ABP57686, ABP57687, ABP57688,
FT      ABP57689, ABP57690, ABP57691, ABP57692, ABP57693, ABP57694,
FT      ABP57695, ABP57696, ABP57697, ABP57698, ABP57699, ABP57700,
FT      ABP57701, ABP57702, ABP57703, ABP57704.
FT      New butenyl-spinosyn biosynthetic genes, useful for increasing the
FT      production of butenyl-spinosyn insecticidal macroliides, or for changing
FT      the metabolites or products produced by spinosyn-producing
FT      microorganisms.
FT      Claim 2; Page 99-119; 218pp; English.
XX      The invention relates to a novel DNA molecule comprising a DNA sequence
XX      that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
XX      PKS (polyketide synthetase) domain, or a spinosyn PKS module. The
XX      butenyl-spinosyn biosynthetic genes are useful for increasing the
XX      production of butenyl-spinosyn insecticidal macroliides. The genes are
XX      also useful for changing the metabolites or products produced by
XX      spinosyn-producing microorganisms. The present sequence represents a DNA
XX      molecule encoding butenyl-spinosyn biosynthetic enzymes.
XX      Sequence 36538 BP; 6867 A; 12266 C; 11182 G; 6223 T; 0 other;
SQ      Query Match
SQ      Best Local Similarity 9.5%; Score 119.6; DB 25; Length 36538;
SQ      Matches 424; Conservative 0; Mismatches 429; Indels 27; Gaps 2;
QY      291 GCTGCCGCGGAGATCAAGCTCGGTCGCAAGCTGTCACCCGCTTTACGTCACG 350
QY      24626 GATGACCCCGCGGAGATCAAGCTCGGTCGCAAGCTGTCACCCGCTTTACGTCACG 24567
QY      351 CGCATTGACCTGCTCGGCGCGGCAATATACAGGCGACCGTTCACCGCTCTGATGCTCG 410
QY      24566 TCGAGTTCGACCAAGCTTCGCGCGCGGACCGAGGCTGTTGACCACTTATTCGAT 24507
QY      411 CTCGGAGACAGAGAGTTGACGTTGCGGGATTACCGGAGGAGATCCGATCGCGCG 470

```

Db 24506 GACCCGAGACGGCGCGCTGGGACCTGGTTCACCCCTGGCCCTGCGCCATCA 24447
 QY 471 GATCAGCGCTCTGTGAAGGTTCCGCGCAGTGTGAGAGAAATTCCGTGCTTGCGCTC 530
 Db 24446 GGTGATCTGCCACTTGTCTGGGATTCCTGGTGGCCAGCCGCGCTTCCGGGTCTGGTC 24387
 QY 531 GCGCATCGCGCGCGCTCGCGGTGGGATTTGGTGGCCAGGTGATGAGAGAACCAAGAC 590
 Db 24386 GGACATTCGTTGGGATCAATGACAGCTCTTAGGAGATTGGGAAAGCCGGATCA 24327
 QY 591 CTGTGCTGCTCCGTTCACCGAGGGGCTGCGCTCTCATGACGTTCTCGATGAGCGCG 650
 Db 24326 AATTGCGCGTATATCGGAGAACTGTCC-----AGCAGCG 24291
 QY 651 CAGGAAACCGGCTCGAAAATGACGTTGACGATGCTGCTCAAGCGCGAGCGGACG 710
 Db 24290 CAAAGAAATGCGCAGTGAAGACTGTCTAGCGTGTGGTTCAGGCGCAAGCGGAGAGC 24231
 QY 711 CAGGCTGAGCAGAGAGAGTGTGCGCTCGTGGGTGCGATTATCGCTGTCGACCGA 770
 Db 24230 CCAGCTGAGCAGAGAGAGATTGTGTGACCGCGCGGGTCTGTGATTCGCGGGTTGA 24171
 QY 771 TACCAAGATCTACTTATCGCGTTCGTGTCTCAACCTGTGCGGTGCGCCGAGCGCT 830
 Db 24170 GACCAACCGCTAACACATCGCAAACTTCACTTCAACCTGTCAACCCATCGGACCACT 24111
 QY 831 CGAGCTGTGAAGCGCGCGCGCTCATAGAGAACCGCTCGATGAGTGTCTCCGCTT 890
 Db 24110 TGACAAAGCTGATCGCGATTCAGAGCTGTGTCCGCGGGTGGAGAGTTGTCTGCGCTA 24051
 QY 891 CGACAATATCTCT---CAGAAATGAGAACTGTGCGTTCGCCAGGCGAGACTGAACTGTG 947
 Db 24050 CACGCGGTGTGGTGCACACCTGGGTTCGCCGATTCGCGACTGAGATCTCGAATCGG 23991
 QY 948 CGGGGCTGATCAAGAAAGGGAGATGTTCTTCTCTGATCCGAGCGCTTGAGAGA 1007
 Db 23990 CGGAGTGTGATTCGCCGGGGTGAAGCAGTGTCTTTCAGATAGCTCGGCAATCTGA 23931
 QY 1008 TGGGACTGTATTTTCAGAGCCAGAGTGTGATGTGGAGCGGAGCAGGGCGGAGCT 1067
 Db 23930 CAGTGCCTCTTTCAGCGGCGGAGCAAACTGACTGTGCGTGGGAGCACAATCTGCAT 23871
 QY 1068 CGCGTACGTAGAGGCCCAATGTCTGCCCGGGGTGTCTCTGCTCGCTCGAGCGGA 1127
 Db 23870 GGGCGTGGGGGACGGCCCGCACTACTGCAATGGGCCCAATTGGGAGATGAGACTCA 23811
 QY 1128 GATGCGGTGGGACCAATCTTCCGTAGGTTCCCGGAGATG 1167
 Db 23810 GGTAGCGATCGGACGCTGATCAACGGTTTCCGAGCTG 23771

RESULT 14

ABN88911 standard; DNA; 1449 BP.

ID ABN88911 standard; DNA; 1449 BP.
 AC ABN88911;
 DT 21-AUG-2002 (first entry)
 DE Streptomyces sp. cytochrome P450 encoding DNA SEQ ID NO:3.
 XX Streptomyces sp. cytochrome P450; anticancer; UCN-01; monooxygenase;
 KM microorganism; gene; ds.
 OS Streptomyces sp.
 XX
 XX Key Location/Qualifiers
 FT CDS 1..1449
 FT /tag= a
 FT /product= "cytochrome P450"
 PN JP2002058490-A.

XX 26-FEB-2002.
 PD 22-AUG-2000; 2000JP-0251304.
 XX 22-AUG-2000; 2000JP-0251304.
 PR 22-AUG-2000; 2000JP-0251304.
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 PA WPI: 2002-410015/44.
 DR P-P8DB; ABN81316.
 PT Novel cytochrome P450 used for improving the productivity of anticancer
 PT drug UCN-01
 PS Claim 5; Page 12-14; 22pp; Japanese.
 CC The present invention describes 3 Streptomyces cytochrome P450 (I)
 CC sequences, ABN81315 to ABN81317 encoded by ABN88910 to ABN88912
 CC respectively. (I) have monooxygenase activity and improve the
 CC productivity of an anticancer drug designated UCN-01. (I) is used
 CC for the production of UCN-01 with transformed Streptomyces sp.
 CC microorganisms.
 SQ Sequence 1449 BP; 231 A; 620 C; 390 G; 208 T; 0 other;

Query March 9.4%; Score 118.6; DB 24; Length 1449;
 Best Local Similarity 48.5%; Pred. No. 9.3e-18;
 Matches 417; Conservative 0; Mismatches 409; Indels 33; Gaps 2;

QY 305 ATCAAGCTGGGATCGGAGCTCGTCAACCCGTGTTTACGTCAAGCGCCATTCGACTGC 364
 Db 521 AACAGCGCCGCTACGCGGCGCTGTGACGAAGGGCTTACACAGCGGGCGGTGGCGGTC 580
 QY 365 TCGCGCGCAATACAGCGCACCGTCAACAGCTGCTGATGCTGCTCCGAGCAAGAG 424
 Db 581 TCGCCCTTACATCTCTCTCTGTCGACGAGTTGCTCGACACCTGGCCGACCAAGGAA 640
 QY 425 AGTTCAAGCTTGTGCGGGATTAGCGGAGGAATCCGATGCGGCGCATAGCGCTGTGT 484
 Db 641 CGGTGACCTGATGCCGACTCGCGGGTCCGCTTCCGTACGATCATGTGCAAGCTGC 700
 QY 485 TGAAGTTCGCGCGAGTGTGACGAGAAGTTCGCTGCTTCCGCTCGGCGACTGCGCGC 544
 Db 701 TCGGGGTGCGGATTC-----CGACCGGCGGTC 728
 QY 545 CGCTCGCGGTGGTTTGGTGTCCAGGTGATGAGAGACCAAGACCTGTGCGGTTCG 604
 Db 729 CGTACGACCTGTGTGAGCG-ACTGTTCGCTGCGGAGACCCGACGCGATCGACGCG 787
 QY 605 TCACGAGAGGGCTCGCGCTGCTCATGACGTTCTTCGATGAGCGGCGCAGAAACCGCTCG 664
 Db 788 CTTCCACAGCGGTGGGACTATGACCGGCTTGTGTGCGGCAAGCGACCGGACCGG 847
 QY 665 AAATGAGTCTTACGATGCTGCTTACAGCGCGAGCGGAGCGAGAGCTGAGACGA 724
 Db 848 GCGACAGCTCTGAGAGACTTATCGCGCTGATGAGCGGAGAGAGACCACTGTCCAGG 907
 QY 725 AGAGCTGTGCGCGCTGTGGGTGCGATTATGCTGTGAGCAGGATCAACGATCTTACC 784
 Db 908 ACGAACTGTATCGCTGCGGCTACTCTGTGTGGCGGCGCAGAGACCACTGCGCAACT 967
 QY 785 TTATCGGTTGCTGTGCTCAACTGTGCGGTGCGGCGGAGCGCTGAGAGTGTGAGG 844
 Db 968 TCATCGGCAAGCGCGCTGCGCTCTTACGCGACCGGAGTGTCTGCGGACCTGAGGG 1027
 QY 845 CCGAGCCGAGCTCATAGAGAACGCGCTCGATGAGAGTCTCCGTTGCAATAATCTCTCA 904
 Db 1028 CCGAGCGCAACTCTCTGCGGCGGCACTGAGACGAGTTGCTGCTTACGACTGCGCGCTG 1087
 QY 905 GAATGAGAACTGTGCTTTTCCGAGGAGACCTGAGATAGTGTGCGGAGCATGATCAAG 964
 Db 1088 GGATCGGACGTTCCGCTTCAGCACGAGGGGCTTACGCTGCGGCGACCGAGATTCGCCG 1147

OY 965 AAGGGAGATGATGCTTTCTTCCTGATCCCGAGGCGCTTGAGAGATCGGACTGATTTCTCA 1024
 Db 1148 AAGGCGTACCGGTCTCTCATTCGACCCCGGCGCGCCCAATCGGACCCGACCGGCTTCCCCG 1207
 OY 1025 GCGCAGACGTGTGTGATGTGCGACGGGACACAGGGGCGGAGCGCTGCGTACGTTAGAGGCC 1084
 Db 1208 ACCCGGACCGGCTGGACCTTCACCCGCGGCGCACCGGTTCACCTGGCTTCCGGCACCGCA 1267
 OY 1085 CCGATGTTGCGCCCGGGGTGTCCCTGTGCTGCGTCCGAGGAGATGCGCGTGGGGCACCA 1144
 Db 1268 TCACCGCTGCTGGAGACCCCGTTGGCCGAGCGCAAGCCGAACTGGCGCTTCCACGCGC 1327
 OY 1145 TCTTCGGTAGGTTCCCGCA 1163
 Db 1328 TCATCACCCTACCCCA 1346
 RESULT 15
 ABX56043
 ID ABX56043 standard, DNA, 1209 BP.
 XX
 AC ABX56043;
 DT 13-FEB-2003 (first entry)
 XX
 XX M. echinospora calicheamicin biosynthesis gene cals.
 KW Calicheamicin biosynthetic gene cluster; aryltetrasaccharide;
 KW aglycone; calicheamicin resistance; nonchromoprotein enediyne;
 KW enediyne resistance; bone marrow cell; gene therapy; gene; ds.
 OS Micromonospora echinospora spp. calichensis.
 FN WO200279465-A2.
 PD 10-OCT-2002.
 PF 28-NOV-2001; 2001WO-US44285.
 PR 28-NOV-2000; 2000US-0724797.
 PA (SLOAN) SLOAN KETTERING INST CANCER *RES.
 PI Thorson J;
 DR WPI; 2003-092897/08.
 DR P-PSDB; ABU11981.
 PT Novel nucleic acid molecule from nonchromoprotein enediyne biosynthetic
 PT gene cluster from Micromonospora echinospora useful for conferring
 PT calicheamicin resistance on a subject -
 PS Claim 9; Page 110-111; 179pp; English.
 CC The present invention relates to the isolation of the Micromonospora
 CC echinospora spp. calichensis calicheamicin biosynthetic gene cluster
 CC encoding proteins and enzymes used in the biosynthetic production of
 CC calicheamicin, including aryltetrasaccharide and aglycone. The gene
 CC cluster also includes the gene encoding for the protein that confers
 CC calicheamicin resistance. The calicheamicin biosynthetic gene cluster
 CC is a nonchromoprotein enediyne biosynthetic gene cluster. Expression
 CC vectors containing genes from the biosynthetic gene cluster are also
 CC disclosed. The expression vectors are useful for producing calicheamicin
 CC biosynthetic proteins. The calicheamicin self-resistance gene provides
 CC an approach for gene therapy, for example, by introduction of enediyne
 CC resistance genes into bone marrow cells, thus increasing resistance and
 CC allowing tolerance to chemotherapeutic doses of calicheamicin.
 CC ABX56026-ABX56073 represent genes from the M. echinospora calicheamicin
 CC biosynthesis gene cluster.
 XX
 XX Sequence 1209 BP; 132 A; 451 C; 452 G; 174 T; 0 other;

Query Match	Similarity	9.3%	Score 116.8	DB 25	Length 1209
Best Local	Similarity 47.7%	Pred. No. 2.3e-17			
Matches 488	Conservative 0	Mismatches 512	Indels 24	Gaps 4	
Qy	128	ACTGAGATGAAAGCGCGCTCTGAGTCTCTCAACCCGATACACAGAGTGTGGCGGTTC	187		
Db	101	ACCGTGGCGCCACCGGGCTGTGATAGTCTTCGCCGCTTCGCGAGTGTGGCGGTCTCG	160		
Qy	188	GCGACGAAACGCTTGGCGGTCAGTGTGAAGAAGTGTGGAAATGTAGCGCGGAATCTCTCG	247		
Db	161	GCGACGAGCGCTTGCCTGGACCGGCATCCGCGGTTCTGAACTGACCTGTGTGGGCGC	220		
Qy	248	CCATTCCCGAGCTCAGCGATATGAAGAATGATGTTGTGGGCTGCCCGGAGATC	307		
Db	221	GGCT-----GCTTCCGAGATCTCGGCGCATCTCTGTTTCCAGACGAGCCGAC	274		
Qy	308	ACGCTCGAGTCCGAGACTGTCAACCCGTCGTTATGTCACCGCGCCATGACCTGTGC	367		
Db	275	ACGGCCCGCTGTGGGGGTGTGGCCCGGGGTTCTGCGGTTCGGCTGCGCGGTGG	334		
Qy	368	GCGCGAAATACAGCGCACGCTGCACCACTGCTCGATGCTTCGTTCCGACAGAGAGT	427		
Db	335	AACCGGTATCGCCGGGACCGTGGACACCTGTGCGCGCGCTTCGCGCGCGGCGCA	394		
Qy	428	TGACGTTTGTGCGGGAATTAAGCGGAGGGAATCCGATGGCGCGGATAGGCGTCTGTGA	487		
Db	395	TGACGTGTGTGACGACCTGGCTGATCCCGCTGGCGGTGTGGCGGCTCTGCGCTGTCG	454		
Qy	488	AGGTTCCGGCGGAGTGTGACGAGAAAGTTCGCTGTTCGCTCGGCGACTGTGCGCGCG	547		
Db	455	GCTGTCCCGCGCGGACTGTGGGGGCGGTTCGGGCGCT-----GGTTCGGCGACG	502		
Qy	548	TGCGCGT	607		
Db	503	TGGGACGGGACCTGTGACCGGGGCGCCAGCCCGGAGGACATGTGCGCGCGCACGCGCGA	562		
Qy	608	CCGAGGGGCTTCGCGCTCTCAATGACGTCTCTCAATGAGCGCGCGAGAACCCGCTGAAA	667		
Db	563	TGCGCGAGTTGTGCGACTGACGTGAGGCGGCGCTTCGAGAGCGCGGCGTGTGAGGCGCG	622		
Qy	668	ATGACGATCTTGACGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	727		
Db	623	AGAGACTGT	679		
Qy	728	AGCTGT	787		
Db	680	AGATGT	739		
Qy	788	TGCGGTTTGT	847		
Db	740	TGGGCAACCGCGT	799		
Qy	848	AGCCCGGAGCTCATGAGGAACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	907		
Db	800	GCGCGGAGCTGT	859		
Qy	908	TAGGAACTGT	967		
Db	860	CCAACACCGGCGAGCTGT	916		
Qy	968	GCGAGATGT	1027		
Db	917	ACGACGT	976		
Qy	1028	CAGACGT	1087		
Db	977	CGGACGATTTGT	1036		
Qy	1088	ATGT	1147		
Db	1037	GCTACTGT	1096		
Qy	1148	TTCG 1151			

Db 1097 CCGG 1100

Search completed: October 3, 2003, 12:28:58
Job time : 317.318 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:41 ; Search time 4507.41 seconds

(without alignments)
11435.853 Million cell updates/sec

Title: US-10-014-717-1_COPY_62369_63628

Perfect score: 1260

Sequence: 1 atgacacagagcagcgaa.....agccctccaagctgcatag 1260

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pac.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pac.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_gy.*
39: em_hlgo_hum.*
40: em_hlgo_mus.*
41: em_hlgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1260	100.0	16124	6 AX024277	AX024277 Sequence
2	1260	100.0	68750	6 AF210843	AF210843 Sorangium
3	1260	100.0	68750	6 AR193029	AR193029 Sequence
4	1260	100.0	68750	6 AR199551	AR199551 Sequence
5	1260	100.0	68750	6 AR199559	AR199559 Sequence
6	1260	100.0	68750	6 AR199567	AR199567 Sequence
7	1260	100.0	68750	6 AR201097	AR201097 Sequence
8	1260	100.0	68750	6 AR208671	AR208671 Sequence
9	1246.2	98.9	58733	1 AF217189	AF217189 Sorangium
10	1246.2	98.9	71989	6 AR172664	AR172664 Sequence
11	157.4	12.5	302650	1 AP005958	AP005958 Bradyrhiz
12	154	12.2	3078	1 SERCP450A	M83110 Saccharopol
13	149.4	11.9	1233	6 AX697983	AX697983 Sequence
14	149.4	11.9	60196	6 AX697977	AX697977 Sequence
15	149.2	11.8	66808	1 SNU421825	AJ421825 Stigmatel
16	148.8	11.8	301925	1 AP005046	AP005046 Streptomy
17	146.4	11.6	78210	1 AB070949	AB070949 Streptomy
18	146.4	11.6	300425	1 AP005022	AP005022 Streptomy
19	143.8	11.4	1257	6 AX097457	AX097457 Sequence
20	143.8	11.4	14759	1 AE004755	AE004755 Pseudomon
21	141.8	11.3	10167	1 SFU08223	U08223 Streptomyc
22	132	10.5	24268	1 AB070947	AB070947 Streptomy
23	132	10.5	300550	1 AP005030	AP005030 Streptomy
24	131.6	10.4	12420	1 AF521896S2	AF521895 Streptomy
25	130	10.3	2243	1 SERERYFGH	M54983 Saccharopol
26	128.8	10.2	1347	6 AX573690	AX573690 Sequence
27	128.8	10.2	90348	1 AF497482	AF497482 Micromono
28	127	10.1	298550	1 AP005029	AP005029 Streptomy
29	123.6	9.8	17512	1 AB071405	AB071405 Lechevali
30	123.6	9.8	25681	1 SAE414559	AJ14559 Saccharot
31	123.6	9.8	26144	1 AB090952	AB090952 Lechevali
32	123.6	9.8	28654	1 AF534707	AF534707 Lechevali
33	123.4	9.8	325483	1 AP005050	AP005050 Streptomy
34	123	9.8	1470	1 AF087022	AF087022 Streptomy
35	123	9.8	4342	1 AF079139	AF079139 Streptomy
36	123	9.8	5970	6 AR271682	AR271682 Sequence
37	123	9.8	5970	6 AR277666	AR277666 Sequence
38	123	9.8	32634	1 AF293355	AF293355 Streptomy
39	122.6	9.7	1191	6 BD133544	BD133544 Novel cyt
40	121.8	9.7	299925	1 AP005039	AP005039 Streptomy
41	121.2	9.6	69644	1 AY179507	AY179507 Streptomy
42	120.6	9.6	210614	1 AB088224	AB088224 Streptomy
43	120.4	9.6	6629	1 AB072568	AB072568 Streptomy
44	120.4	9.6	19601	1 AB070944	AB070944 Streptomy
45	120.4	9.6	299425	1 AP005049	AP005049 Streptomy

ALIGNMENTS

RESULT 1
LOCUS AX024277 16124 bp DNA
DEFINITION Sequence 82 from Patent DE19846493.
ACCESSION AX024277 AX024384
VERSION AX024277.1 GI:10184551
KEYWORDS
SOURCE
ORGANISM
Polyangium cellulosum
Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 Beyer, S. and Mueller, R.J.
AUTHORS Patent: DE 19846493-A 82 13-APR-2000;
JOURNAL BIOTECHNOLOG FORSCHUNG GMBH (DE)

COMMENT On Oct 15, 2002 this sequence version replaced gi:10184588.
FEATURES
source
1.16124
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/db_xref="taxon:56"
BASE COUNT 2336 a 5400 c 5695 g 2693 t
ORIGIN
Query Match 100.0%; Score 1260; DB 6; Length 16124;
Best Local Similarity 100.0%; Pred. No. 2.3e-199;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGACACAGAGACAGACGATGAGTGAAGACGCTTTCGACTTCAAGCCGTTG 60
Db 7086 ATGACACAGAGACAGACGATGAGTGAAGACGCTTTCGACTTCAAGCCGTTG 7145
Qy 61 GGGCCCTGGGTACGCGGAGAGACCGGTTCCCGCGATGAGCGGCTGAGAGAGGCAACCCG 120
Db 7146 GGGCCCTGGGTACGCGGAGAGACCGGTTCCCGCGATGAGCGGCTGAGAGAGGCAACCCG 7205.
Qy 121 ATCTTCTAGGATGAGAGCGGCTCTTGGGTCCTTCAACCGATACACAGCGTGTGAGG 180
Db 7206 ATCTTCTAGGATGAGAGCGGCTCTTGGGTCCTTCAACCGATACACAGCGTGTGAGG 7265
Qy 181 GTGTTCCGCGAGACAGCGCTTCCGCGTCACTCCAGAAAGATGGGAATCGACCGCGACTAC 240
Db 7266 GTGTTCCGCGAGACAGCGCTTCCGCGTCACTCCAGAAAGATGGGAATCGACCGCGACTAC 7325
Qy 241 TCGTCGCGCATTCGCGAGCTCAGCGATGATGAGAAATGTTTCGCGGCTGCGCGCG 300
Db 7326 TCGTCGCGCATTCGCGAGCTCAGCGATGATGAGAAATGTTTCGCGGCTGCGCGCG 7385
Qy 301 GAGGATCAGCTTCGCGGTCGCGAAGCTGTCACACCCGCTTTCAGTACGCGCGCATGAC 360
Db 7386 GAGGATCAGCTTCGCGGTCGCGAAGCTGTCACACCCGCTTTCAGTACGCGCGCATGAC 7445
Qy 361 CTGCTCGCGCGGCAATATACGCGGACCGTTCGACAGCTGCTGATCTGCTCGGACAA 420
Db 7446 CTGCTCGCGCGGCAATATACGCGGACCGTTCGACAGCTGCTGATCTGCTCGGACAA 7505
Qy 421 GAGGATTCGAGCTTTCGCGGATTCACGCGGAGGAAATCCCGATGCGCGCATCAGCGCT 480
Db 7506 GAGGATTCGAGCTTTCGCGGATTCACGCGGAGGAAATCCCGATGCGCGCATCAGCGCT 7565
Qy 481 CTGTTGAAGGTTTCGCGCGAGTGTGACGAGAAATTCGCTTCGCGCTCGCGCATGCG 540
Db 7566 CTGTTGAAGGTTTCGCGCGAGTGTGACGAGAAATTCGCTTCGCGCTCGCGCATGCG 7625
Qy 541 CGCGCGCTCGCGGCTGCTTGGTTCGCGGACGCTGATGAGAGACCAAGACCTGCTGCGG 600
Db 7626 CGCGCGCTCGCGGCTGCTTGGTTCGCGGACGCTGATGAGAGACCAAGACCTGCTGCGG 7685
Qy 601 TCGCTACCGAGGAGGCTCGGCTGCTCATGACGTCCTCGATGAGCGCGCAGAGAACCG 660
Db 7686 TCGCTACCGAGGAGGCTCGGCTGCTCATGACGTCCTCGATGAGCGCGCAGAGAACCG 7745
Qy 661 CTGGAATAAGCTTCTGACGATGCTTTCAGGCGGAGCGCAGCGACGCTGAGC 720
Db 7746 CTGGAATAAGCTTCTGACGATGCTTTCAGGCGGAGCGCAGCGACGCTGAGC 7805
Qy 721 AGGAAGAGAGCTGTCGCGCTCGGCTGAGTATATGCTGCTGCGACCGATACCAAGATC 780
Db 7806 AGGAAGAGAGCTGTCGCGCTCGGCTGAGTATATGCTGCTGCGACCGATACCAAGATC 7865
Qy 781 TACCTATGCGCTTCTGCTGCTCATGACGTCGCGTCCGCGAGCGCTCGAGCTGATG 840
Db 7866 TACCTATGCGCTTCTGCTGCTCATGACGTCGCGTCCGCGAGCGCTCGAGCTGATG 7925
Qy 841 AAGGCGAGCGCGGCTCATGAGAAACGCGCTGATGAGGTGCTTCGCTTCAAGATATC 900
Db 7926 AAGGCGAGCGCGGCTCATGAGAAACGCGCTGATGAGGTGCTTCGCTTCAAGATATC 7985

Qy 901 CTCGATATGAGAACTGTCGTTTTCGCGGAGGACGCTGAGTACTGCGGGCATCGATC 960
Db 7986 CTCGATATGAGAACTGTCGTTTTCGCGGAGGACGCTGAGTACTGCGGGCATCGATC 8045
Qy 961 AAGAAAGGAGATGCTCTTCTCTCTGATCCGAGCGGCTGAGAGATGGAATCTGATTC 1020
Db 8046 AAGAAAGGAGATGCTCTTCTCTCTGATCCGAGCGGCTGAGAGATGGAATCTGATTC 8105
Qy 1021 TCCAGCGCAGACGCTTTCGATGTCGACGCGACACGCGCGGAGCGCTTCCGATGGA 1080
Db 8106 TCCAGCGCAGACGCTTTCGATGTCGACGCGACACGCGCGGAGCGCTTCCGATGGA 8165
Qy 1081 GGGCGCCATGTCGCGCGGGGTGTCCTTGTGCTGCTGAGGCGGAGATGCGCGTGGG 1140
Db 8166 GGGCGCCATGTCGCGCGGGGTGTCCTTGTGCTGCTGAGGCGGAGATGCGCGTGGG 8225
Qy 1141 ACCATCTTCGATGCTTCCCGAGATGAGAGCTGAAAGAACTCCCGTTCGATGACAC 1200
Db 8226 ACCATCTTCGATGCTTCCCGAGATGAGAGCTGAAAGAACTCCCGTTCGATGACAC 8285
Qy 1201 CCGCGCTTCGAGACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATG 1260
Db 8286 CCGCGCTTCGAGACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATG 8345
RESULT 2
AF210843 68750 bp DNA linear BCT 21-JAN-2000
LOCUS
DEFINITION
Sorangium cellulosum strain so ce90 epothione biosynthesis gene cluster, complete sequence.
ACCESSION
AF210843
VERSION
AF210843.1 GI:6724237
KEYWORDS
ORGANISM
Polyangium cellulosum
Polyangium cellulosum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE
1 (bases 1 to 68750)
Molnar, I., Schnupp, T., Ono, M., Zirkle, R., Milmanow, M., Nowak-Thompson, B., Engel, N., Toupet, C., Stratmann, A., Cyr, D. D., Gorlach, J., Mayo, J. M., Hu, A., Goff, S., Schmid, J. and Ligon, J. M. The biosynthetic gene cluster for the microtubule-stabilizing agents epothiones A and B from Sorangium cellulosum so ce90 Chem. Biol. 7 (2), 97-109 (2000)
20130945
MEDLINE
PUBMED
2 10662695
REFERENCE
Molnar, I.
TITLE
JOURNAL
Submitted (03-DEC-1999) Natural Product Genetics, Novartis Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O. Box 12557, Research Triangle Park, NC 27709, USA
FEATURES
source
1.68750
/organism="Polyangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce90"
/db_xref="taxon:56"
<1.1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AA26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSAVLDDGHRATERYHLAEARGIEDIALREH
LRIOGSPFCHMCGLDLELLADQPLAISI FPHASLRHPDTPSAMVYDGLV
RMLARGAPGLREYEEERBARATQERLMLAAAPCPADLPFRPDNDANGLPG
MSPEVAERLRASATYRPELACALLAMLTGAGPWSGAYAEMLPENLILGRLPT
ALASAPGTSIALRGARLPASWEVSSKKSQSLGIPALMEHLRTIYPMANGADN
LSRPERAIAEVRRLRAQAPAPFAGSLAVAGVSSGRLSGVLTDGALYSGDND
IVMOPGRISPVLLAGTDPEFELAPLSOMLFVHANAGTISKVLTEGSPILIVARN
QARPWSLVHARGFMVNWQAMVPDERGAFVVGQRSTIMEFEHPTRCLHBPAGSAPS

CDS
LACDEHLWCELSAGRLLEWRHHPHRRGAPRAYIGENHIAATWYPSLTNATHVL
WADRRALILADKRTGVEPIVLAETIRPPHVVSEDRDIFALTQGPSRDMVHEHIR
SGASTVADYOROLMDRPDVLINRGLEFFTNDRILLTARS"
complement (1900..3171)
/note="ORF2; similar to hypothetical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to D0-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAAPCALGAEERQPGQDAGALAPAREVMAEVA
AGOMGAWMLVARGDVDVAVTELGASAPMRDITFIASMTKATVAVMVLVE
EGKLDSPVDRLMPELANRKVLARIGPIDETVPABRPITVRDLMFTMGFGISPLA
SSPIORADEGLVNAOPVPMTPHGPDEWIRLITGLPLMHOPGAMNTYTSYQVIL
VGRADQCFDAPVREIRIAPGMRDIDPHVADKLAPAGCGFTDEGTETMDHD
GAESVAPSPAPEGAGLVSTVDYLLFARMNGVHEGRLLSAASVREMTADHL
TPAKAASSFPPEFETHGMVGMAVVAVPDPAVSEVPGRYGMDGFGTSMINDGRELI
IGIWMTSAGFLFSGALERFWRSYVATESA"
3415..5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGTEBOVLSTVTLITVTAASGELARLRQPVIGELRG
GVVIGPSVGLAPGFHRLAQEPBAVGVVLSIGISVIGLMLNLAAGLEVDGILRKEA
RPGALSALGAIAPLAPAGASALVLDPLSGILGLVLSLVAVSAKVLIERESM
RSLYAQVTLAGVSEVAVMVLVAMTSSYSGASALVARSALASGLITPMVYVGR
LTLAMRVADATRVSKGVSILVLTFLAALQORLAPHLGLAFALGVILNAGRT
NRPLDGVQTLVAGLFAPVFVLAMRVDSQLTPAMVGALLATATAKVTPAA
LGALGCRGSEALVAVGLMKGTDLIVAVVEGLLSNEATYAVVAVATVTRA
SPALLMLERAPPTQESARLERBEAKRAYITGEVRIIVIVAHALPGATUVES
IVASKRKLGETVDITELSEQAPGSPRAAGSGLARLAVGIMRORRLRGS
IOAILRASRDLDITIGARSPARAGMSFGLQDAIVQRAESNVLVVGDDEPAERAS
ARLITVPIIGESFAADLAHVALAMDALVELLSSAQTDPGAVWMDREPSVRAV
ARSLVDEAVFGRRLGVSVSSRVHVGARPSDEITRELARAYDILLVIGCYHGLRL
YLGSVSVSVRSVPVALLVYHAGTREQVR"
complement (5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGTRCAIAEGYIPGSGNGPEPQMTSHETACLINASDRDA
QVATVFSDDPDAGPYVTPARTTRHVRNDLTPEPPIPRDITDVASVISEDPIVY
OHTRLDSROAENALLSTIAYTDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHYDTGRRRGRRIQHTLGLLASALLAGCGSPSEKTYQGTRELAP
GADARVADVDPAATYTRLADVAVHLSPEERLEAGSERFVWQRPSPSPRRRGVLD
YNADSRGKLAEITVPYANFELLITAEKQSSPQSPSSAAVIGPISVG"
7610..11875
/gene="epoa"
7610..11875
/gene="epoa"
/note="Epos A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRIERAAEDPIAIVGASCRLEPGVVIDLSGEWTLLEGSRDITV
GRVAPERWDAAMFDPDPAKGTPTVTRASTSDVACPDASFPGISPEALRMDPAHR
LLLEVCWALENAIAIPASLVGTETGVFIGIGPSEYEAALPQATASAEIDAHGILGTM

PSVAGRISVALGRPCVAVDTAYSSSLVAVHILACQSLRSGECSTALAGVSLMLSP
STVIAHSEKTRALADGCKAFSAEDAFDGGEGCAVVVLRSLSGARADGRILAVIG
SAIWHDSASGLITVNGSSOEIVLKRALADAGCAASVAVLBSGATDGLTPIEIOA
LNAVYGRVAPVAPPLIGSVKTNLGHVEVSGITGLVILSHOGQIPATHIOALN
PRISWGLRITVTRAPTPWDMTPRAGVSSRMSGTNNHVVIEEPPAATCTPABE
RPAELVLASRTASALDOAARLDHLETPSOQLGVAASLATTREAMERLVAAVT
SREGLRALDAAGQOTSPGAVRSIAOSSKGLAFLTQGAQTLGMRGILYDWSAF
REAFDLCVRLFNOELDRLEBVMAEASVDALDQTAQTPALPFEFEALALMS
MGVEPELVAGHSIGELVACVAGVFSIVIAAGPPIVAAARGMALPAGAVSIEAPPA
DYAAAVAPHAASVIAAVNAPDOVVIAGAOCPVHAIANAARARAKALHVSAPHS
PLVAPMLAEFARVAVESVYRPSVILVNSLGACTQEVSPGIVWRHAEVVPFAG
VKALHAAGCTPVEVGRKSTLGLVPCMDPARALLASRAGDEBATAVLEALGGM
AVGGLVSMAGLPFSGGRVVDLPTYPMGREYVIDTKADDAAGDRAPAGHDEBEG
GAVGGRSARABDHPPESSGRREKVEAAGRPRLTIDEPGDLHLVYTERARPC
LGEVEIVDAAGSLSDNFVOLALGVNPDLPKGPAPPLIGGECGRIVAAGEVNGVY
VGQPVIALSAGATHTYTSAAVLPERPQALSAEAAAMPAYATANYALDRILRQ
GERVLIIAATGGVGLAVVQAQHVGAHVATAGPEKALVLESIGVYDSRSDRFV
ADVAVATGGGVGLVNLNSLSELIDKSFNLLRSHGRVELEKDCVADNQLRPLR
NLSPSLVDLGMMLNREPARVRLLEELGLIAGVFTPPATLPARVADAFSMQ
AOHLGKLVLTGDEPQIRIPTHAGAPSTGDRDLRLASAPARAAALAEFLRTQ
VSOVLARTPEIKVGAELFTRLGMSLMAVEILRNIEASLKLKSLTFLSTSPNIALLA
ONILDLATATLSLERVAENLRAGVQNDPFSGADQWEIITAL"
7640..8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epop"
11872..16104
/gene="epop"
/note="EPOS P; contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINOLINLEHOGIKLADGERLQIQAPNALPNLLARISEH
KSTILMLRQLPAESIVPAPAEERHAFPLTDIOESYWLGTGAFVPSGIAHAREYD
CTDLDVRLSRAPFKVVARHDMLRHMLPMMQVIEBKVDADIEILRGLDRSTREA
RLVSLDRAMSHRIYDTERPPLVYVAVNRDLERQRLVLSIDLINVDLSGSLTFKDM
SYEDPETSLEPESLRYDYVLALDESKSEBAHORSMDYMRRIAELEPPETLPMKAD
PSTLKEIRFRHTEQMLPDSWGRLLKRVGRGLTPVTVILAAFEVILGWSASPERFTL
NITLFNRLPVHPRVNDITGDTSMVLLDITTRKSEFOAKRIOEOLWEMADCDVS
GIVEORAAVVLGIORGALPEVVLTLNQOVGVSTLORGTAVYSTOTPOLLDH
OLYVEDGDVLAMDIYDGVPPDLDDMLSLAAVYVFLRLREPEGBOVSGSLPAOLE
ABASANATMALISHTLHGLFAARVQLPMOLAVSARLITVEELSRSRRLGARLR
EQGARPTLVAVWMEKHEBOVAVLALLEGAAVYIPADILPARIRIYLLDHGKYL
LTPQRLDGKLSWPEIQRLVSEAGVAGDQPPMP1QTPSDLAIVYISGSGLPK
GVMIDHGAANTLLIDINERPEIGDGVNLASLSPLSYDVGILAAAGTIVPDA
SKLADPAHMAELIEREKVTWNSVPALMRMLVEFEGRPLSILASLSDMIPV
GLPELOAIRGVSVISLGGATASINSIGVAVNVLSMSIPIYGRPLNRQTHVD
GALERPVMVPGOLYIGVGALGALGMDKTRKSPVHPETGERLKYTGDLGRVLD
GNIEFGRBNQIKLRGYRVLGEIETLASHPRVNAVYVPGVNDANULLAVYVP
EGTRRRAEDDASLKTERRIDARAAHADGSLDSERVOFKLARGLRDLGKRVLD
TQODPREAGLDVYARRSRVTFLEAPLPVEFGFTLSKLSVEDGATLPKFRPVSJ
STYVQTYAVVKSRIEGVDEGFYYYPFPHRLILKSDSHIGSCINQVROGDFDEEA
FNLLFVGRIDAISLIGSSRBEFLAGVMAOLMDAASGNCINQVROGDFDEEA
PVLDLRSDVYVHGMCGRVDPDROFOVCTIGCSPPRATTRGAPRERDHFAMLD
FLTRKLEPVYPTVFEVLDLPLTSNKRVRKALREKQDTPSHSGHTPRDLEEL
LVAAVREVLAEVGLQOSFVDIGATSIIHVRMSLLOKRLDREIATTELFOYPNLGS
LASGLRDSDDLQDPRMORVREVRKGRRRS"
16251..21749
gene

CDS
/gene="epob"
16251..21749
/gene="epob"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AF26920.1"

Query Match 100.0%; Score 1260; DB 1; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.8e-199;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGACACAGAGACCAACCAATCAGATGAGACGACACCTGCTTCACTTCAAGCCGTTG 60
DB 62369 ATGACACAGAGACCAACCAATCAGATGAGACGACACCTGCTTCACTTCAAGCCGTTG 62428
QY 61 GCGCCTGGGTACCGGAGACCCGTTCCCGCATGAGCGCCTGAGAGGCAACCCG 120
DB 62429 GCGCCTGGGTACCGGAGACCCGTTCCCGCATGAGCGCCTGAGAGGCAACCCG 62488
QY 121 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGTCCTGACCCGATACCAAGCCTGTGAGG 180
DB 62489 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGTCCTGACCCGATACCAAGCCTGTGAGG 62548
QY 181 GTGTTCCGCGACGACAGCTTCCGCGTCACTGCGAAGAGTGGGAATCGACCGCGAGTAC 240
DB 62549 GTGTTCCGCGACGACAGCTTCCGCGTCACTGCGAAGAGTGGGAATCGACCGCGAGTAC 62608
QY 241 TCGTCGGCCATTCCGAGCTCAGCGATATGAAAGATGAGATGTTGTCGGGCTGCGCGG 300
DB 62609 TCGTCGGCCATTCCGAGCTCAGCGATATGAAAGATGAGATGTTGTCGGGCTGCGCGG 62668
QY 301 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGTCGTTTACGTACCGCGCATGAC 360
DB 62669 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGTCGTTTACGTACCGCGCATGAC 62728
QY 361 CTGCTCGCGCGCGCAATATACAGCGCACCGTGCACCACTGCTCGATGCTGCTCGGACAA 420
DB 62729 CTGCTCGCGCGCGCAATATACAGCGCACCGTGCACCACTGCTCGATGCTGCTCGGACAA 62788
QY 421 GAGGATTCGACGTTGTCGGGATTAACGCGAGGAGATCCGATGCGCGCATCAGCGCT 480
DB 62789 GAGGATTCGACGTTGTCGGGATTAACGCGAGGAGATCCGATGCGCGCATCAGCGCT 62848
QY 481 CTGTTGAGGTTCCGCGCGCAAGTGTGACGAGAGTTCGTCGCTTCCGCGACTGCG 540
DB 62849 CTGTTGAGGTTCCGCGCGCAAGTGTGACGAGAGTTCGTCGCTTCCGCGACTGCG 62908
QY 541 CCGGCGCTCGCGTGGGTTTGTGCGCCGAGTGCAGAGAGACCAAGACCCGTCGCGG 600
DB 62909 CCGGCGCTCGCGTGGGTTTGTGCGCCGAGTGCAGAGAGACCAAGACCCGTCGCGG 62968
QY 601 TCGTCACCGAGGCGCTCGCGCTGCTCAGTACGTCCTCGATGAGCGCGCAGAGAACCG 660
DB 62969 TCGTCACCGAGGCGCTCGCGCTGCTCAGTACGTCCTCGATGAGCGCGCAGAGAACCG 63028
QY 661 CTGGAATAAGCTCTTGAAGATGCTGCTTCAAGCGGAGCGCAGCGCAGCGCTGAGC 720
DB 63029 CTGGAATAAGCTCTTGAAGATGCTGCTTCAAGCGGAGCGCAGCGCAGCGCTGAGC 63088
QY 721 AGCAAGAGCTGCTCGCGCTCGGAGTGTGAGTATAGCTGCTGCGACCCATTAACCAAGATC 780
DB 63089 AGCAAGAGCTGCTCGCGCTCGGAGTGTGAGTATAGCTGCTGCGACCCATTAACCAAGATC 63148
QY 781 TACCTTATCGCGTTCGCTGCTCAACTGCTCGGTCGCGCCAGAGCGCTCGAGCTGATG 840
DB 63149 TACCTTATCGCGTTCGCTGCTCAACTGCTCGGTCGCGCCAGAGCGCTCGAGCTGATG 63208
QY 841 AAGGCGAGCCGCGGCTCATGAGAAACGCGCTGATGAGGTGCTCGGCTTTCGACATATC 900
DB 63209 AAGGCGAGCCGCGGCTCATGAGAAACGCGCTGATGAGGTGCTCGGCTTTCGACATATC 63268

```

```

QY 901 CTCAGATATGAACTGTGCTTTTCGCGAGGACGACCTGAGTACTGCGGGCATGCATC 960
DB 63269 CTCAGATATGAACTGTGCTTTTCGCGAGGACGACCTGAGTACTGCGGGCATGCATC 63328
QY 961 AAGAAAGGAGATGCTTTCTCTCTGATGCCGAGCGCTGAGAGATGGAATGTATTC 1020
DB 63329 AAGAAAGGAGATGCTTTCTCTCTGATGCCGAGCGCTGAGAGATGGAATGTATTC 63388
QY 1021 TCCAGCCGACGCTGTTGATGTCGACGCGGACACCGGCGGAGCTCCGTTACGCTTGA 1080
DB 63389 TCCAGCCGACGCTGTTGATGTCGACGCGGACACCGGCGGAGCTCCGTTACGCTTGA 63448
QY 1081 GGGCCCAATGCTGCGCGGGGTGTCCTGTGCTGCTCGAGCGGAGATCGCGTGAGC 1140
DB 63449 GGGCCCAATGCTGCGCGGGGTGTCCTGTGCTGCTCGAGCGGAGATCGCGTGAGC 63508
QY 1141 ACCATCTTCCGTAGGTTCCCGAGATGAAAGTGAAGAACTCCGCTTTGATATACAC 1200
DB 63509 ACCATCTTCCGTAGGTTCCCGAGATGAAAGTGAAGAACTCCGCTTTGATATACAC 63568
QY 1201 CCGCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 1260
DB 63569 CCGCGCTTCGGAACATGATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 63628

```

RESULT 3
AR193029
LOCUS AR193029 68750 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6346404.
ACCESSION AR193029
VERSION AR193029.1 GI:20238994
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schupp, T., Ligon, J., Madison, I., Molnar, I., Zirkle, R., Cyr, D., Dawn, and
Goriach, J.
TITLE Genes for the biosynthesis of epothenones
JOURNAL Patent: US 6346404-A 1 12-FEB-2002;
FEATURES
source location/Qualifiers
1..68750
/organism="unknown"

BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 1260; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.8e-199;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGACACAGAGACCAACCAATCAGATGAGACGACACCTGCTTCACTTCAAGCCGTTG 60
DB 62369 ATGACACAGAGACCAACCAATCAGATGAGACGACACCTGCTTCACTTCAAGCCGTTG 62428
QY 61 GCGCCTGGGTACCGGAGACCCGTTCCCGCATGAGCGCCTGAGAGGCAACCCG 120
DB 62429 GCGCCTGGGTACCGGAGACCCGTTCCCGCATGAGCGCCTGAGAGGCAACCCG 62488
QY 121 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGTCCTGACCCGATACCAAGCCTGTGAGG 180
DB 62489 ATCTTCTAGTGGATGAGAGCGGCTCTGAGGTCCTGACCCGATACCAAGCCTGTGAGG 62548
QY 181 GTGTTCCGCGACGACAGCTTCCGCGTCACTGCGAAGAGTGGGAATCGACCGCGAGTAC 240
DB 62549 GTGTTCCGCGACGACAGCTTCCGCGTCACTGCGAAGAGTGGGAATCGACCGCGAGTAC 62608
QY 241 TCGTCGGCCATTCCGAGCTCAGCGATATGAAAGATGAGATGTTGTCGGGCTGCGCGG 300
DB 62609 TCGTCGGCCATTCCGAGCTCAGCGATATGAAAGATGAGATGTTGTCGGGCTGCGCGG 62668
QY 301 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGTCGTTTACGTCAACGCGCATGAC 360
DB 62669 GAGGATCACGCTCGGGTCCGCAAGCTGTCACCCGTCGTTTACGTCAACGCGCATGAC 62728

```


QY 361 CTGCTGCGCGCGGAATAACAGCCACCGTCAGACAGCTGCTGATGCTGCTCGCCGAGAA 420
Db 62729 CTGCTGCGCGCGGAATAACAGCCACCGTCAGACAGCTGCTGATGCTGCTCGCCGAGAA 62788
QY 421 GAGGAGTTGAGCTGTGTGCGGGATTACCGCGAGGGAAATCCGATGCGCGCATGACCGCT 480
Db 62789 GAGGAGTTGAGCTGTGTGCGGGATTACCGCGAGGGAAATCCGATGCGCGCATGACCGCT 62848
QY 481 CTGTTGAAGGTTCCGGCGGAGTGTGACGAGGAAGTTCCGTCGCTTGCGGCTCGCGCATGCG 540
Db 62849 CTGTTGAAGGTTCCGGCGGAGTGTGACGAGGAAGTTCCGTCGCTTGCGGCTCGCGCATGCG 62908
QY 541 CGCGGCTCGCGGCTGTGTGTGTCGCCAGTGTGATGAGAGACCAAGACCTGTGTGCG 600
Db 62909 CGCGGCTCGCGGCTGTGTGTGTCGCCAGTGTGATGAGAGACCAAGACCTGTGTGCG 62968
QY 601 TCCGTCACCGAGGGGCTGCGCTGCTCCATGACGTCCTGATGAGCGCGCGAGAACCG 660
Db 62969 TCCGTCACCGAGGGGCTGCGCTGCTCCATGACGTCCTGATGAGCGCGCGAGAACCG 63028
QY 661 CTCGAAATGACGTCCTGATGAGTGTGCTGAGCGCGGCGAGCGCGAGCGCTGAGCTGAGT 720
Db 63029 CTCGAAATGACGTCCTGATGAGTGTGCTGAGCGCGGCGAGCGCGAGCGCTGAGCTGAGT 63088
QY 721 ACGAAGAGCTGTGCTGCGCTCGTGGGTGAGATTATGCTGCTGCGACCGATACACAGATC 780
Db 63089 ACGAAGAGCTGTGCTGCGCTCGTGGGTGAGATTATGCTGCTGCGACCGATACACAGATC 63148
QY 781 TACCTTATCGCTGTGTGCTGCTCAACCTGCTGCGGTGCGCGAGCGCTGAGCTGAGT 840
Db 63149 TACCTTATCGCTGTGTGCTGCTCAACCTGCTGCGGTGCGCGAGCGCTGAGCTGAGT 63208
QY 841 AAGCGCGAGCGCGGCTCATGAGAAACGCGCTCATGAGTGTGCTGCTGCAAAATATC 900
Db 63209 AAGCGCGAGCGCGGCTCATGAGAAACGCGCTCATGAGTGTGCTGCTGCAAAATATC 63268
QY 901 CTCAGAAATGAGAACTGTGCTGCTGCGCGAGGACGCTGAGTATGCTGCGGCGCATGATC 960
Db 63269 CTCAGAAATGAGAACTGTGCTGCTGCGCGAGGACGCTGAGTATGCTGCGGCGCATGATC 63328
QY 961 AAGAAAGGAGAGTGTGCTGCTGCTGATCCGAGCGCGCTGAGAGATGGAATGCTGATTC 1020
Db 63329 AAGAAAGGAGAGTGTGCTGCTGCTGATCCGAGCGCGCTGAGAGATGGAATGCTGATTC 63388
QY 1021 TCCAGGCGAGAGTGTGATGATGAGAGCGGACAGCGGCGCGAGCTGCGTGAACGCTAGA 1080
Db 63389 TCCAGGCGAGAGTGTGATGATGAGAGCGGACAGCGGCGCGAGCTGCGTGAACGCTAGA 63448
QY 1081 GCGCGCGATGCTGCGCGGAGTGTGCTGCTGCTGAGCGCTGAGAGGAGATGCGCGTGAAC 1140
Db 63449 GCGCGCGATGCTGCGCGGAGTGTGCTGCTGCTGAGCGCTGAGAGGAGATGCGCGTGAAC 63508
QY 1141 ACCATCTTCCGATGATTCGCCGAGATGAGCTGAAAGAACTCCGCTGTTTGATATACAC 1200
Db 63509 ACCATCTTCCGATGATTCGCCGAGATGAGCTGAAAGAACTCCGCTGTTTGATATACAC 63568
QY 1201 CCGCGCTTCCGAGAAATGAAATCACTCAACGTCATCTTGAAGCCTCCCAAACTGGAATG 1260
Db 63569 CCGCGCTTCCGAGAAATGAAATCACTCAACGTCATCTTGAAGCCTCCCAAACTGGAATG 63628

RESULT 4
AR199551 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR199551 Sequence 1 from patent US 635457.
ACCESSION AR199551
VERSION AR199551.1 GI:20249625
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)

AUTHORS Schupp, T., Iigou, J., Madison, J., Molnar, I., Zirkle, R., Cyr, D., Dawn, and
Gorlach, J.
TITLE Genes for the biosynthesis of epoethionones
JOURNAL Patent: US 635457-A 1 12-MAR-2002;
FEATURES Location/Qualifiers
SOURCE 1. 68750
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 100.0%; Score 1260; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.8e-199; Mismatches 0; Indels 0; Gaps 0;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGACACAGAGAGCAAGCAATCAGATGAGAGCAAGCTGCTTTCAGCTTCAAGCCGTTG 60
Db 62369 ATGACACAGAGAGCAAGCAATCAGATGAGAGCAAGCTGCTTTCAGCTTCAAGCCGTTG 62428
QY 61 GCGCTGGGTACCGGAGAGACCCGTTCCCGCATGAGCGCTGAGAGAGCAACCCCG 120
Db 62429 GCGCTGGGTACCGGAGAGACCCGTTCCCGCATGAGCGCTGAGAGAGCAACCCCG 62488
QY 121 ATCTTCTACTGGATGAGAGCCGCTCTGCGGTCTCTACACCGCATACAGAGCTGTGCGG 180
Db 62489 ATCTTCTACTGGATGAGAGCCGCTCTGCGGTCTCTACACCGCATACAGAGCTGTGCGG 62548
QY 181 GTGTTCCGCGAGAGACGCTTGCGCGGTGAGTGAAGAGTGGAAATCGAGCGCGAGTAC 240
Db 62549 GTGTTCCGCGAGAGACGCTTGCGCGGTGAGTGAAGAGTGGAAATCGAGCGCGAGTAC 62608
QY 241 TCGTCGCGCAATTCGCGAGCTCAGCATATGAGAAATGAGATGTTGCGGCTGCGCGG 300
Db 62609 TCGTCGCGCAATTCGCGAGCTCAGCATATGAGAAATGAGATGTTGCGGCTGCGCGG 62668
QY 301 GAGGATCACTGCTGCGGTCCGCAAGCTCTCAACCCGCTGTTTACGTCAAGCGCCATCGAC 360
Db 62669 GAGGATCACTGCTGCGGTCCGCAAGCTCTCAACCCGCTGTTTACGTCAAGCGCCATCGAC 62728
QY 361 CTGCTGCGCGCGGAATTCAGCGGACGCTGACACGAGCTGCTGATGCTGCTGCGCGAGAA 420
Db 62729 CTGCTGCGCGCGGAATTCAGCGGACGCTGACACGAGCTGCTGATGCTGCTGCGCGAGAA 62788
QY 421 GAGGAGTTGCACTGTGTGCGGATTAACGCGAGGAGATCCGATGCGCGCATCAGCGCT 480
Db 62789 GAGGAGTTGCACTGTGTGCGGATTAACGCGAGGAGATCCGATGCGCGCATCAGCGCT 62848
QY 481 CTGTTGAAGGTTCCGGCGGAGTGTGACGAGAAATTCGCTGCTGCGCTGCGCGCATGCG 540
Db 62849 CTGTTGAAGGTTCCGGCGGAGTGTGACGAGAAATTCGCTGCTGCGCTGCGCGCATGCG 62908
QY 541 GCGCGGCTCGCGGCTGTGTGTGTCGCCAGTGTGATGAGAGACCAAGACCTGTGTGCG 600
Db 62909 GCGCGGCTCGCGGCTGTGTGTGTCGCCAGTGTGATGAGAGACCAAGACCTGTGTGCG 62968
QY 601 TCCGTCACCGAGGGGCTGCGCTGCTCCATGACGTCCTGATGAGCGCGCGAGAACCG 660
Db 62969 TCCGTCACCGAGGGGCTGCGCTGCTCCATGACGTCCTGATGAGCGCGCGAGAACCG 63028
QY 661 CTCGAAATGACGTCCTGATGAGTGTGCTGAGCGCGGCGAGCGCGAGCGCTGAGCTGAGT 720
Db 63029 CTCGAAATGACGTCCTGATGAGTGTGCTGAGCGCGGCGAGCGCGAGCGCTGAGCTGAGT 63088
QY 721 ACGAAGAGCTGTGCTGCGCTGCTGAGCGGATGAGCGCGCGAGCGCGAGCGCTGAGCTGAGT 780
Db 63089 ACGAAGAGCTGTGCTGCGCTGCTGAGCGGATGAGCGCGCGAGCGCGAGCGCTGAGCTGAGT 63148
QY 781 TACCTTATCGGCTGTGCTGCTCAACCTGCTGCGGTGCGCGGCGCTGAGCTGAGT 840
Db 63149 TACCTTATCGGCTGTGCTGCTCAACCTGCTGCGGTGCGCGGCGCTGAGCTGAGT 63208
QY 841 AAGCGCGAGCGCGGCTCATGAGAAACGCGCTCATGAGTGTGCTGCTGCAAAATATC 900
Db 63209 AAGCGCGAGCGCGGCTCATGAGAAACGCGCTCATGAGTGTGCTGCTGCAAAATATC 63268

QY 901 CTCAGAAATGAACTGTGCTTTCGCCAGGCAAGCACTGAGTACTGCGGGGATGATC 960
DB 63269 CTCAGAAATGAACTGTGCTTTCGCCAGGCAAGCACTGAGTACTGCGGGGATGATC 63328
QY 961 AAGAAAGGGAGATGTGTTCTTCTCTGATCCGAGGCGCTGAGAGATGGAGTATTC 1020
DB 63329 AAGAAAGGGAGATGTGTTCTTCTCTGATCCGAGGCGCTGAGAGATGGAGTATTC 63388
QY 1021 TCCAGGCCAGAGCTGTTTGAATGTCGACGAGGCGAGGCTGCGTACGATGAC 1080
DB 63389 TCCAGGCCAGAGCTGTTTGAATGTCGACGAGGCGAGGCTGCGTACGATGAC 63448
QY 1081 GGGCCCATGTCTGCCCCGGGGGTGTCCTTGTCTGCTGAGGCGAGATTCGCGTGGG 1140
DB 63449 GGGCCCATGTCTGCCCCGGGGGTGTCCTTGTCTGCTGAGGCGAGATTCGCGTGGG 63508
QY 1141 ACCATCTTCGTAAGTTCGCCGAGATGAAGCTGAAGAAACTCCGTTTGGATACAC 1200
DB 63509 ACCATCTTCGTAAGTTCGCCGAGATGAAGCTGAAGAAACTCCGTTTGGATACAC 63568
QY 1201 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 1260
DB 63569 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 63628

RESULT 5
ARI99559 68750 bp DNA linear PAT 20-APR-2002
LOCUS ARI99559 Sequence 1 from patent US 6355458.
DEFINITION ARI99559
ACCESSION ARI99559
VERSION ARI99559.1 GI:20249633
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355458-A 12-MAR-2002;
FEATURES
source 1. 68750
location/Qualifiers
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 1260; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.8e-199;
Matches 1260; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACACAGAGAGCAAGCAATCAGATGAGACGAGCCTGCTTTCGACTTCAAGCCGTTT 60
DB 62369 ATGACACAGAGAGCAAGCAATCAGATGAGACGAGCCTGCTTTCGACTTCAAGCCGTTT 62428
QY 61 GGGCCCTGGGTAAGCGGAGGAGCCGTTTCCCGGATATGAGGCGCTTGAGAGAGGAAACCCC 120
DB 62429 GGGCCCTGGGTAAGCGGAGGAGCCGTTTCCCGGATATGAGGCGCTTGAGAGAGGAAACCCC 62488
QY 121 ATCTTCTACGTGAGTAAGGCGCTCTGAGGTCCTCAACCGATACACAGAGTGTGGGG 180
DB 62489 ATCTTCTACGTGAGTAAGGCGCTCTGAGGTCCTCAACCGATACACAGAGTGTGGGG 62548
QY 181 GTGTTCCGCGACGACGCTTTCGCGTCACTGTCGAGAAAGTGGGAATCGACGCGAGTAC 240
DB 62549 GTGTTCCGCGACGACGCTTTCGCGTCACTGTCGAGAAAGTGGGAATCGACGCGAGTAC 62608
QY 241 TCGTCCGCAATTCGCGAGCTCAGCGATATGAAAGATGAGATGTTTGGGCTGGCCGCG 300
DB 62609 TCGTCCGCAATTCGCGAGCTCAGCGATATGAAAGATGAGATGTTTGGGCTGGCCGCG 62668
QY 301 GAGGATACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTACGCGCCATCGAC 360

DB 62669 GAGGATACGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTACGCGCCATCGAC 62728
QY 361 CTGCTGCGGCGGAAATACAGGCAACCGTCCAGCAAGCTGCTGATGCTGCTCCGAGCAA 420
DB 62729 CTGCTGCGGCGGAAATACAGGCAACCGTCCAGCAAGCTGCTGATGCTGCTCCGAGCAA 62788
QY 421 GAGGATTCGACGTTGTGCGGGATTAAGCGAGGAGATCCCGATGCGGCGATCAGCGT 480
DB 62789 GAGGATTCGACGTTGTGCGGGATTAAGCGAGGAGATCCCGATGCGGCGATCAGCGT 62848
QY 481 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAGTTCCGTGCTTGGGCTCGGCGACTGCG 540
DB 62849 CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAGTTCCGTGCTTGGGCTCGGCGACTGCG 62908
QY 541 CGCGGCTGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600
DB 62909 CGCGGCTGCGGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 62968
QY 601 TCCGTCACCGAGGGGCTGCGGCTGCTCGATGACGCTCGATGAGCGGGCGAGAACCG 660
DB 62969 TCCGTCACCGAGGGGCTGCGGCTGCTCGATGACGCTCGATGAGCGGGCGAGAACCG 63028
QY 661 CTCGAAATGACGTCCTTGAACGATGCTGCTTCAAGCCGAGGCGGAGCGGACGAGCTGAGC 720
DB 63029 CTCGAAATGACGTCCTTGAACGATGCTGCTTCAAGCCGAGGCGGAGCGGACGAGCTGAGC 63088
QY 721 ACGAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780
DB 63089 ACGAAGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63148
QY 781 TACCTTATCGCGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 63149 TACCTTATCGCGCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63208
QY 841 AAGGCCGAGCCCGGCTCATGAGGAAAGCGCTCATGAGTGTCTCGCTTGACAAATATC 900
DB 63209 AAGGCCGAGCCCGGCTCATGAGGAAAGCGCTCATGAGTGTCTCGCTTGACAAATATC 63268
QY 901 CTCAGAAATGAACTGTGCTTTCGCCAGGCAAGCACTGAGTACTGCGGGGATGATC 960
DB 63269 CTCAGAAATGAACTGTGCTTTCGCCAGGCAAGCACTGAGTACTGCGGGGATGATC 63328
QY 961 AAGAAAGGGAGATGTGTTCTTCTCTGATCCGAGGCGCTGAGAGATGGAGTATTC 1020
DB 63329 AAGAAAGGGAGATGTGTTCTTCTCTGATCCGAGGCGCTGAGAGATGGAGTATTC 63388
QY 1021 TCCAGGCCAGAGCTGTTTGAATGTCGACGAGGCGAGGCTGCGTACGATGAC 1080
DB 63389 TCCAGGCCAGAGCTGTTTGAATGTCGACGAGGCGAGGCTGCGTACGATGAC 63448
QY 1081 GGGCCCATGTCTGCCCCGGGGGTGTCCTTGTCTGCTGAGGCGAGATTCGCGTGGG 1140
DB 63449 GGGCCCATGTCTGCCCCGGGGGTGTCCTTGTCTGCTGAGGCGAGATTCGCGTGGG 63508
QY 1141 ACCATCTTCGTAAGTTCGCCGAGATGAAGCTGAAGAAACTCCGTTTGGATACAC 1200
DB 63509 ACCATCTTCGTAAGTTCGCCGAGATGAAGCTGAAGAAACTCCGTTTGGATACAC 63568
QY 1201 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 1260
DB 63569 CCCGCGTTCGGAACATCGAATCACTCAACGTCATCTTGAAGCCCTCCAAAGCTGATAG 63628

RESULT 6
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS ARI99567 Sequence 1 from patent US 6355459.
DEFINITION ARI99567
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6355459-A 12-MAR-2002;
FEATURES
source 1. 68750
location/Qualifiers
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Db	62669	GAGGATACAGCTCGGGGTCCCGAAGCTGTGAACCCGTCGTTTACGTACAGCGGCATCGAC	62728
Oy	361	CTGCTGGCGCCGGAATACAGCGCACCTGTGACCAAGCTGTCGATGCTGTGCTCCGACAA	420
Db	62729	CTGCTGGCGCGCCGGAATACAGCGCACCGTTCGACCAAGCTGTCGATGCTGTGCTCCGACAA	62788
Oy	421	GAGGAGTTTCGACGTTTGTCGGGGAATTACGGGAGGGAAATCCCGATGCGCGGATTCAGCGCT	480
Db	62789	GAGGAGTTTCGACGTTTGTCGGGGAATTACGGGAGGGAAATCCCGATGCGCGCGGATTCAGCGCT	62848
Oy	481	CTGTTGAAGGTTCCGGCCGAGTGTGACGAAAGTTCGTCGTCTCGCTCGCGCATGTCG	540
Db	62849	CTGTTGAAGGTTCCGGCCGAGTGTGACGAAAGTTCGTCGTCTCGCTCGCGCATGTCG	62908
Oy	541	CGCGCGCTCGCGCGTGGGTTTGTCGTCGCCAGCTCGATGAGGACCAAGACCTTGTCGCG	600
Db	62909	CGCGCGCTCGCGCGTGGGTTTGTCGTCGCCAGCTCGATGAGGACCAAGACCTTGTCGCG	62968
Oy	601	TCCGTCACCGAGGGGCTCGCGCTGTCGATGACGTCCTCATGAGCGCGCGCAGGAAACCG	660
Db	62969	TCCGTCACCGAGGGGCTCGCGCTGTCGATGACGTCCTCATGAGCGCGCGCAGGAAACCG	63028
Oy	661	CTCGAAAATGACGTCCTTGAAGATGCTGCTTCAAGCGCAGCGCAGCGCAGCGAGGCTGAGC	720
Db	63029	CTCGAAAATGACGTCCTTGAAGATGCTGCTTCAAGCGCAGCGCAGCGCAGCGAGGCTGAGC	63088
Oy	721	ACGAGGAGCTGATCGCGCTCTGTGGGTGCGATTAATGCTGTCGACCGAATACAGATC	780
Db	63089	ACGAGGAGCTGATCGCGCTCTGTGGGTGCGATTAATGCTGTCGACCGAATACAGATC	63148
Oy	781	TACCTTATCGGTTTCGCTGTGCTCAACCTGCTGCGGTTCGCCAGGCGCTCCAGCTGGTG	840
Db	63149	TACCTTATCGGTTTCGCTGTGCTCAACCTGCTGCGGTTCGCCAGGCGCTCCAGCTGGTG	63208
Oy	841	AAGGCCGAGCCCGGGGCTCATGAGGAAAGCGCTCGATGAGGTGCTCCGCTTTCGACAAATATC	900
Db	63209	AAGGCCGAGCCCGGGGCTCATGAGGAAAGCGCTCGATGAGGTGCTCCGCTTTCGACAAATATC	63268
Oy	901	CTCAGAAATAGGAACCTGTGCGCTTTCGCCAGGACAGGACCTGAGATGATCGGGGATGATC	960
Db	63269	CTCAGAAATAGGAACCTGTGTGCTTTCGCCAGGACAGGACCTGAGATGATCGGGGATGATC	63328
Oy	961	AAGGAAAGGGGAGATGTCCTTTCTCTGATATCCCGAGGCGCTTGAGAGATGGGACTGTATTC	1020
Db	63329	AAGGAAAGGGGAGATGTCCTTTCTCTGATATCCCGAGGCGCTTGAGAGATGGGACTGTATTC	63388
Oy	1021	TCCAGGGCCAGACGATTTGATGTGTCAGCGGAGACACGGGCGCGAGCTCGGATACGGTATA	1080
Db	63389	TCCAGGGCCAGACGATTTGATGTGTCAGCGGAGACACGGGCGCGAGCTCGGATACGGTATA	63448
Oy	1081	GGCCCCCATGTCGCCCCCGGGGTGTCCTTTCCTGCTCGAGGCGGAGATTCGCGGTGGC	1140
Db	63449	GGCCCCCATGTCGCCCCCGGGGTGTCCTTTCCTGCTCGAGGCGGAGATTCGCGGTGGC	63508
Oy	1141	ACCATTTTCGTAGGTTCCCGAGATGAAGCTGAAGAAATCTCCGCTTTTGATATCCAC	1200
Db	63509	ACCATTTTCGTAGGTTCCCGAGATGAAGCTGAAGAAATCTCCGCTTTTGATATCCAC	63568
Oy	1201	CCCGGTTTCGGGAACATCGAATCATCTAAGGTATCTGTAAGGCGCTCCAAAGCTGGATAG	1260
Db	63569	CCCGGTTTCGGGAACATCGAATCATCTAAGGTATCTGTAAGGCGCTCCAAAGCTGGATAG	63628

RESULT	8
LOCUS	AR208671
DEFINITION	Sequence from patent US 6383787.
ACCESSION	AR208671
VERSION	AR208671.1
KEYWORDS	GI:2150986
SOURCE	.
ORGANISM	Unknown.
	Unknown.

REFERENCE	Unclassified.	1 (bases 1 to 68750)
AUTHORS	Schupp, T., Ligon, J., Madison, J., Molnar, I., Zirkle, R., Cyr, D., Dawn, and Gorlach, J.	
TITLE	Genes for the biosynthesis of epothilones	
JOURNAL	Patent: US 6383787-A 1 07-MAY-2002;	
FEATURES	Location/Qualifiers	
source	1..68750	
	/organism="unknown"	
BASE COUNT	9596 a 22456 c 25539 g 11159 t	
ORIGIN		
Query Match	100.0%; Score 1260; DB 6; Length 68750;	
Best Local Similarity	100.0%; Pred. No. 1,8e-199;	
Matches 1260; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
1	ATGACACAGAGGACGGAATCAGATGAGACGAGCCCTGCTTTCGACTTCAAGCCGTTT	60
62369	ATGACACAGAGGACGGAATCAGATGAGACGAGCCCTGCTTTCGACTTCAAGCCGTTT	62428
61	GCGCTTGGGTTACCGCGAGGACCCGTTTCCCCGCAATGACGCGCTGAGAGAGGCAACCCC	120
62429	GCGCTTGGGTTACCGCGAGGACCCGTTTCCCCGCAATGACGCGCTGAGAGAGGCAACCCC	62488
121	ATCTTACTGAGGATGAGAGCGCTCTGAGGATCTTCAACCGATACACACGCTGCGCG	180
62489	ATCTTACTGAGGATGAGAGCGCTCTGAGGATCTTCAACCGATACACACGCTGCGCG	62548
181	GTGTTCCGCGACGACGCTTCGCGCTCAGTCGAGAGAGTGGAGATTCGACCGCGAGTAC	240
62549	GTGTTCCGCGACGACGCTTCGCGCTCAGTCGAGAGAGTGGAGATTCGACCGCGAGTAC	62608
241	TGCTGGCCATTCCCGAGCTCAGCGATATGAGAGATACGATTTGTTCCGCGTCCCGCG	300
62609	TGCTGGCCATTCCCGAGCTCAGCGATATGAGAGATACGATTTGTTCCGCGTCCCGCG	62668
301	GAGATACGCTCGGGTCCGCAAGCTGCTCAACCCGCTTACGTCACGCGCAATCGAC	360
62669	GAGATACGCTCGGGTCCGCAAGCTGCTCAACCCGCTTACGTCACGCGCAATCGAC	62728
361	CTGCTGCGCGCCGAAATACAGCGCACCGTTCGACCTGCTCGATGCTTCCGCGACAA	420
62729	CTGCTGCGCGCCGAAATACAGCGCACCGTTCGACCTGCTCGATGCTTCCGCGACAA	62788
421	GAGAGTTTGAAGCTTGTGCGGATTAACGCGAGGAGATCCGATGCGCGCATCAGCGCT	480
62789	GAGAGTTTGAAGCTTGTGCGGATTAACGCGAGGAGATCCGATGCGCGCATCAGCGCT	62848
481	CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGTGCGCTGCGCGCATGCG	540
62849	CTGTTGAAGGTTCCGCGCGAGTGTGACGAGAAAGTTCCGTGCGCTGCGCGCATGCG	62908
541	CGCGCGCTCGCGCTGGGTTGTGAGCCCGAGTGTGATGAGAGACCAAGACCTTGTCGCG	600
62909	CGCGCGCTCGCGCTGGGTTGTGAGCCCGAGTGTGATGAGAGACCAAGACCTTGTCGCG	62968
601	TCCGTCAACGAGGAGGCTCGCGCTGCTCATTGACGCTCTTCAATGAGCGGCGACGAA	660
62969	TCCGTCAACGAGGAGGCTCGCGCTGCTCATTGACGCTCTTCAATGAGCGGCGACGAA	63028
661	CTGCAAAATACGCTTTGACAGATGCTGCTTACAGGCGGAGGCGGACGCGGCGGCGG	720
63029	CTGCAAAATACGCTTTGACAGATGCTGCTTACAGGCGGAGGCGGACGCGGCGGCGG	63088
721	ACGAGAGAGCTGGTTCGCGCTCGTGGGTCGATTAATGCTGCTGCGCACCGATACGATC	780
63089	ACGAGAGAGCTGGTTCGCGCTCGTGGGTCGATTAATGCTGCTGCGCACCGATACGATC	63148
781	TACCTTATCGCGTTCGCTGCTCAACCTGCTGCGGCTCGCGGCGGCTGAGCTGGTG	840
63149	TACCTTATCGCGTTCGCTGCTCAACCTGCTGCGGCTCGCGGCGGCTGAGCTGGTG	63208
841	AGGCGGAGCCCGGCTCATAGAGAAACGCGCTGATGAGGTGCTCCGCTTCCAGCAATATC	900

[illegible]


```
0. YEHDDGLVLAAMDIVDVPDPDLDDMLAEVAVFLRLTTEPMSEOMKCSUPLAOLE
AASANETMSLSETHLHGLPAARPEQLPMQALAVSARKTITYEELSRSRRLGARLR
EGGAPNTLVAVVMEKMEQVAVVAVLESQAIVPIDDLPALERTHYHLLDGEVXLV
LQPMWDLGSLWPGIQRLLVSDAVGEGDGLPMWIIQTPSDLAIVYITLSDGVLK
GVMIDHRAVANTLIDINERFEGPGRVATLSLSDLESVDVGLAAGTIVPDA
SKLRDPHAAVALIEREKTVVMSVPALMELVEHSEGRDLSARSRLSLGSDVPI
GLBGLQAIRPGVSVISIGATEASISWISGYPVNRVPSMASIPYGRPLRNTQTHLD
EALBERRPVPGQULVIGVGLAGTIRDEKTRNSFLVHPERGEIYKTKDGLRPLD
GNIEMGRENDQIKLKGIVELGETEELTKSHPNVADAVIVGANDAKLLAVLP
ECTRRRAAQDQSLKTERVDARAAKADGLSDGERVQKLAHGLRLDLDKPVVDL
TLVPEREAGLDVARRSVRTLEAPIPVEFERFLSCLSVEPDGALPKRRYSAG
STYPOTVYAKSGRLBEGVDEGYYVHPPEHLLKMDSDIERGAHPVNPVPEBA
PGLTIVGRIDATLESYGLSREBCELEAGMYLMEQAPSCNIGCAVPGDPFQVR
PYLDRHSVYVHGHGVRVDPQOVCTLGDSSRRATTTGAPRGROHFDLIRD
PRTKLPEYVPTVFEVLDALPUTSNGKDRALKRKTSSPRSHGHTAPDABEII
LVAVREVLGLEVNGLOSFDLGATSIHIVMRSILQRLDREAIETELFOYPILGS
LASGLRDSKDLRPNMQRVYEARKRGRRS"
10639..16137
/gene="epoc"
10639..16137
/gene="epoc"
/notes="type I polyketide synthase; epochlione PKS module
2"
/codon_start=1
/cranel_table=11
/product="Bpc"
/protein_id="AA62882.1"
/db_xref="GI:7453559"
/translation="MEQESSAIAVIGSGRFGPAGRLDEPMRNLADTTEAVQRFSEBQ
BLASGVDPALVLDPMYVRAGSVLEEDVDFDAFPISSREALMDPORIIMECME
ALEMNGYDPTAYEGSIGVYAGANMSYLSLNSHEHPANMAGITOTILIGNKIYAT
HYSYRLNLBPSISVOTACSTSLVAVHLACMSLIDRECMALAGITVIAIPRACVY
AAGGILSPDHGRAPAKANGITIMNGCCVLLKPLDRLSGDPRAVILDSATND
GARKIGTAPSEVGAQOALMEALALAGVARSIOYIETHTGTLLDALETALRVF
DADASTRRSCAIGVKTGIGLESAGIAGIKTIVALEHRQLPSLINESPNSIDE
ASSPFYNTSLKDMNTSTPRRAGVSSFGIGTNVAVLEAPALPAAPARAEAL
FVYSKSAALADLAAARLDHLOAHQGLSDGVAFLATVSRPMERILMAAREBAL
REGDAARAGOTPRGAVRGRCSPGVPRVYVPRGOSOMVOMQOLABERVPAL
SACDRALQBAKSHLAEALADEGSSQIRIVQVYVPLALVAPALRMSRGPBDV
VIGHSVEVAAAHVAGALSLEDAVALICRSRLRRI SQQGEAAVETLSLAEABLR
GYEDRVSVAVNSPRSTVLSGEPAALIGEVLSSINAGVCRVKVDVSHSPQVPLR
EDLLAALGRLPRAAIVPMRSTVTGAMVGPSELGANNRQPVRFVAVQADLQ
GHGFEVMSPHPLITTSVEEMBRAAQRAAGVLSRGDSEPRAMLEALGALVRLP
VWGRLFRPAGSRRLVPTYPMQREPVYITAPAKSAGDRGRPAGHRLTGMQNTST
QYSTRIMETTLDLKPLMLGDRVQCAVVPFPAAYIEMALISSGAELACGPIQITDV
LAELAFAGDAAVLVQVVTTEQPSGRLOFIASRAAGAHASFRVHARGALRVERTE
VPAGLITSAVRARLQSMBAATYAEITMGLQYEPAFQGIELNREGEDALGRVLP
DAAGSAEYRLPALIDACFOVVGSLFAGGEBATPMVPEVGSRLLOPSEGLCHA
RVNHGROTPDRQAGDFMVVDSGAVVAGVSLVQRLFGVRREEDWPLLEMEP
AAVGTAKVNAAGMILLGGGGGALRSWLEAGHAYVHAASNTSAAAGVRLAKAF
DQCAPTAIVHLSLDGGGELDPGLAGHGLDAPRSADVSPDLDPALVGCDSVLMYV
QALAGWFRDAPRLWLTTRGAOVAGADVSTQAPLLIGRYIAMEHADURCARVLD
PRPDEIGALALAEALADAEVALRGERCVARIVRQPEPRGRJESCVPTDVY
IADSTYLVLTGGIGLGSVAGMLERGAGHLVLRSGAASEQGAVALAEAGAR
VVAAXADVADRAOLERILREVTSGMGLGVAAAGIYLDGLMOOTPRPFXKAP
VOGALHALTRBAPLSFVLYASGVGLGSPGCGYAAANTEFLALHRRBAQGLPA
LEVDWGLPAVEGMAAQBDRGARLVSRKRSILTPDGSLAKRLBESGVQVCNPN
PRUWELYPALASRMLSLVTRHRSAGPAGDGLRLALAEASASGSLLEPLRL
AOISQVRLRBEKIEVDAPLITLGNMSLNGLEIRNLEMLQITVATPLMTYPTVAA
LEGLHARACEAPVSPHTTADSAVEIEEMSDDLTQILAAKFALT"
16134..37907
/gene="epoc"
16134..37907
/gene="epoc"
/notes="type I polyketide synthase; epochlione PKS modules
3, 4, 5, and 6"
/codon_start=1
/cranel_table=11
/product="epod"
/protein_id="AA62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAOQNPLKOAIIIORELSRAGLAAQALERTPIAIV
GIGCRPGADAPBEATFWEILDERDAVQGLDRMVLVGAPVBAVPMHMGILTEPIDC
```


Db 57597 AAGCCGAGCCGGGCTCATGAGGAACGGCTCGATGAGTCTCCGCTTCACAAATATTC 57656
Qy 901 CTCAGATATGGAACGTGTGCTTTCGCGAGGCGAGACCTGGAGTACTGGGGGCATCGATTC 960
Db 57657 CTCAGATATGGAACGTGTGCTTTCGCGAGGCGAGACCTGGAGTACTGGGGGCATCGATTC 57716
Qy 961 AAGAAAGGGAGATGATGCTTCTCTCTGATCCGAGGCGCTGAGATGGAGCTGATATTC 1020
Db 57717 AAGAAAGGGAGATGATGCTTCTCTCTGATCCGAGGCGCTGAGATGGAGCTGATATTC 57776
Qy 1021 TCAGAGCCAGACGTGTGATGTGCGACGGGACACGGGCGCAGAGCTTGGCGTATGCA 1080
Db 57777 TCAGAGCCAGACGTGTGATGTGCGACGGGACACGGGCGCAGAGCTTGGCGTATGCA 57836
Qy 1081 GGGCCCAATGTCTGCGCCGGGGGTGTCCTGTGCTGCTGAGGCGGAGATGCGCGTGGGC 1140
Db 57837 GGGCCCAATGTCTGCGCCGGGGGTGTCCTGTGCTGCTGAGGCGGAGATGCGCGTGGGC 57896
Qy 1141 ACCATCTTCCGTAGGTTCCCGAGATGAGAGTGAAGAACTCCGCTGTTTGATATCAC 1200
Db 57897 ACCATCTTCCGTAGGTTCCCGAGATGAGAGTGAAGAACTCCGCTGTTTGATATCAC 57956
Qy 1201 CCCGCTTCCGGAACATGATCACTCAACGTTCATCTTGAAGCCCTCCAAAGCTGATTA 1259
Db 57957 CCCGCTTCCGGAACATGATCACTCAACGTTCATCTTGAAGCCCTCCAAAGCTGATTA 58015

RESULT 10
ARI72664
LOCUS ARI72664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION ARI72664
VERSION ARI72664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothenones C and D
JOURNAL Patent: US 6303342-A 2 16-OCT-2001;
FEATURES
source 1..71989
/organism="unknown"
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 98.9%; Score 1246.2; DB 6; Length 71989;
Best Local Similarity 99.4%; Pred. No. 3.4e-197;
Matches 1251; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 ATGACACAGAGCAAGCCGATCAAGTGAAGACGAGCCGCTTTCGACTTCAAGCCGTTTC 60
Db 56757 ATGACACAGAGCAAGCCGATCAAGTGAAGACGAGCCGCTTTCGACTTCAAGCCGTTTC 56816
Qy 61 GCGCGTGGGTACGCGGAGGAGCCGTTCCCGCGATCGAGCGGCTGAGAGGAGCAACCCC 120
Db 56817 GCGCGTGGGTACGCGGAGGAGCCGTTCCCGCGATCGAGCGGCTGAGAGGAGCAACCCC 56876
Qy 121 ATCTTCTACGTGGATGAAGGCGCTCTGAGTCTCAACCCGATACACAGCTGTCCGCG 180
Db 56877 ATCTTCTACGTGGATGAAGGCGCTCTGAGTCTCAACCCGATACACAGCTGTCCGCG 56936
Qy 181 GTGTTCCGCGAGCAACGCTTCCGCGTCACTCGAAGAGTGGGAATCGAGCCGAGATAC 240
Db 56937 GTGTTCCGCGAGCAACGCTTCCGCGTCACTCGAAGAGTGGGAATCGAGCCGAGATAC 56996
Qy 241 TCGTGCGGCAATCCCGAGGCTGACGATATGAGAAAGTATGAGTGTGGGCTGCGCGCG 300
Db 56997 TCGTGCGGCAATCCCGAGGCTGACGATATGAGAAAGTATGAGTGTGGGCTGCGCGCG 57056
Qy 301 GAGGATCAAGCTCGGGTCCGCAAGCTCGTCAACCCGTCGTTTACGTACGCGCATCGAC 360

Db 57057 GAGGATCAAGCTCGGGTCCGCAAGCTCGTCAACCCATGTTTACGTACGCGCATTCAC 57116
Qy 361 CTGCGTGGCGCGGAATATACAGGACCGTGCAGACAGCTGCTCGATGCTCGGCTCGAGAA 420
Db 57117 CTGCGTGGCGCGGAATATACAGGACCGTGCAGACAGCTGCTCGATGCTCGGCTCGAGAA 57176
Qy 421 GAGAGTTCGACGTGTGTCGGGATTAACGCGAGGAGATCCCGATGCGCGCATACGCGT 480
Db 57177 GAGAGTTCGACGTGTGTCGGGATTAACGCGAGGAGATCCCGATGCGCGCATACGCGT 57236
Qy 481 CTGTTGAAGTTCGCGCCGAGTGTGACGAGAAATTCCGTCCGTTGCGCTCGGCACTGCG 540
Db 57237 CTGTTGAAGTTCGCGCCGAGTGTGACGAGAAATTCCGTCCGTTGCGCTCGGCACTGCG 57296
Qy 541 GCGCGCTCGCGCGTGGGTTTGTGTGCGCCGAGTGCATGAGAGAACCAACCTGCTCCCG 600
Db 57297 GCGCGCTCGCGCGTGGGTTTGTGTGCGCCGAGTGCATGAGAGAACCAACCTGCTCCCG 57356
Qy 601 TCCGTCAACGAGGGCTCGCGCTGCTCCATGACGTCTCGATGAGCGGCGAGAACCCG 660
Db 57357 TCCGTCAACGAGGGCTCGCGCTGCTCCATGACGTCTCGATGAGCGGCGAGAACCCG 57416
Qy 661 CTCGAAATGACGTCTTGAAGATGCTCTTCAAGCCGAGCCGACGAGCAAGCTGAGC 720
Db 57417 CTCGAAATGACGTCTTGAAGATGCTCTTCAAGCCGAGCCGAGCAAGCTGAGC 57476
Qy 721 ACGAAGAGCTGTGTCGCGCTGCTGGGTCGATTAATCGCTGTGCGACCGATACCAACGATTC 780
Db 57477 ACGAAGAGCTGTGTCGCGCTGCTGGGTCGATTAATCGCTGTGCGACCGATACCAACGATTC 57536
Qy 781 TACCTTATCGGCTTCCGCTGCTCAACCTGCTGCGGTCGCCGAGGCGCTCGAGCTGATG 840
Db 57537 TACCTTATCGGCTTCCGCTGCTCAACCTGCTGCGGTCGCCGAGGCGCTCGAGCTGATG 57596
Qy 841 AAGCGCGAGCCCGGCTCATGAGAAACCGCTCGATGAGTGTCTCCGTTGCAAAATATC 900
Db 57597 AAGCGCGAGCCCGGCTCATGAGAAACCGCTCGATGAGTGTCTCCGTTGCAAAATATC 57656
Qy 901 CTCGAAATGAGAACTGTGCGCTTTCGCGAGGAGACCTGGAAGTACTGGGGGCATCGATTC 960
Db 57657 CTCGAAATGAGAACTGTGCGCTTTCGCGAGGAGACCTGGAAGTACTGGGGGCATCGATTC 57716
Qy 961 AAGAAAGGGAGATGATGTTCTCTCTGATCCGAGCGCTGAGAGATGGAGTATATTC 1020
Db 57717 AAGAAAGGGAGATGATGTTCTCTCTGATCCGAGCGCTGAGAGATGGAGTATATTC 57776
Qy 1021 TCAGAGCCAGACGTGTGATGTGCGACGGGACACAGGCGCGAGCTTGGCGTATGCA 1080
Db 57777 TCAGAGCCAGACGTGTGATGTGCGACGGGACACAGGCGCGAGCTTGGCGTATGCA 57836
Qy 1081 GGGCCCAATGTCTGCGCCGGGGGTGTCCTGTGCTGCTGAGGCGGAGATGCGCGTGGGC 1140
Db 57837 GGGCCCAATGTCTGCGCCGGGGGTGTCCTGTGCTGCTGAGGCGGAGATGCGCGTGGGC 57896
Qy 1141 ACCATCTTCCGTAGGTTCCCGAGATGAGAGTGAAGAACTCCGCTGTTTGATATCAC 1200
Db 57897 ACCATCTTCCGTAGGTTCCCGAGATGAGAGTGAAGAACTCCGCTGTTTGATATCAC 57956
Qy 1201 CCCGCTTCCGGAACATGATCACTCAACGTTCATCTTGAAGCCCTCCAAAGCTGATTA 1259
Db 57957 CCCGCTTCCGGAACATGATCACTCAACGTTCATCTTGAAGCCCTCCAAAGCTGATTA 58015

RESULT 11
AP005958/c
LOCUS AP005958 302650 bp DNA linear BCT 28-MAR-2003
DEFINITION Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 24/31.
ACCESSION AP005958 BA000040
VERSION AP005958.1 GI:27354550
KEYWORDS
SOURCE Bradyrhizobium japonicum USDA 110

ORGANISM	Bradyrhizobium japonicum USDA 110
REFERENCE	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.
AUTHORS	1 Kaneoko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriyuchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110
JOURNAL	DNA Res. 9 (6), 189-197 (2002)
MEDLINE	22484998
PUBMED	12597275
REFERENCE	2 Kaneoko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriyuchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S.
TITLE	Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110 (supplement)
JOURNAL	DNA Res. 9 (6), 225-256 (2002)
MEDLINE	22485002
PUBMED	12597279
REFERENCE	3 (bases 1 to 302650)
AUTHORS	Kaneoko, T.
TITLE	Direct Submission
JOURNAL	Submitted (20-NOV-2002) Takakazu Kaneoko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7
MEDLINE	Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
PUBMED	(E-mail: kaneoko@kazusa.or.jp.
REFERENCE	URL: http://www.kazusa.or.jp/rhizobase/ Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934
FEATURES	Location/Qualifiers
source	1. .302650 /organism="Bradyrhizobium japonicum USDA 110" /mol_type="genomic DNA" /strain="USDA110" /db_xref="taxon:224911" 31. .1110 /gene="blr6272" 31. .1110 /gene="blr6272" /note="AraC family ORF ID:blr6272" /codon_start=1 /transl_table=11 /product="transcriptional regulatory protein" /protein_id="BACS1537.1" /db_xref="GI:27354551"
gene	/translations="MENTLESVESFGLGGLPVASGRDAKGRAMFTDLOAPPAVID LDAEPDHLQLOAOAI PERGWRIRIDRPEAIEDITTSRWEDFSSSDHATTED RYFVGIALKATRAKLTDRQVIFDGTMPAGTLVSPKPLAOFQAPAFHFIISA DHFPASPAEKGIDLVLDPLAELARLATEHGAADREFTRCVQOTLAFHISA PPRARNALPKRIRREVOYIADNPFRCISLSLAVAGISRHHPAAGRTATGVAR EYLINRIRIHAKTILATTGRLAETILAVGFSQAHFSTVFKRISGSAARRRLASKA EPILAEALPRRRPADRDWMAASAA"
CDS	complement(1428. .3794) /gene="b116273" complement(1428. .3794) /gene="b116273" /note="ORF ID:b116273 probable biotin sulfoxide reductase" /codon_start=1 /transl_table=11 /protein_id="BACS1538.1" /db_xref="GI:27354552"
gene	/translations="MDRYTATWYRNKDKMMDDTIGPPDGLDLSDFKPTSHWGVF ARNGAGLEVAYACDPDPNGITDNFPGALRHQARIAQPAIRRGHIERGPGDPRGR DERVSVSEKADLDLDELGRIRDTGPOAVFGSGSAGFHHASQVHRFLAIA MGGYRVSNSVSGSALLPQILAGYEDITKRNVTWEIATETDVLAFGMLKNS MVGAGISIKHVERSAMEARKGCEFTLVSPLEADLPVAGAEWMTCVSTPTALMG IVTTLSEKLDQAFLDRTYEGWPFRLRTGSDGPPVAGAEWMAAISGVADTTITL AARLAKRALITYSHSLQRAHGEQPVMMGMVLAALGLGILPGGVAISLAIQVYG
gene	BRVNDVPGPTLGGGRNGVRDPIPVARIADMLNPGATRYNGETRTYDILRVWAG NPRHHODINLRAKARFADLTFVYVHELAWTAARHADIYESTMTLEDEDIGYSEND LMVAHRIAEFFGLARDYELFADLAERLGRPEPTBERTREMIKHIYETPRASLAM RGLAEPSDFEWRQSLVLPQPDGGLRFRFDVPVHAHLPBGRLEIETSTRJAG GDADCPGPWLDITDMKPGAPCEFLVANOQVTRHLSQLDGSGSLGAKHGREVARM NPVADARGIKDGDILRFENRGACLAHVTDGLAPGVOLPTGSAWYDPMDEDDA LCVGNPVLTRDGTSSFAQCGQLTTEVEKKTGNLPLIRAFDPV"
gene	complement(3898. .4947) /gene="b116274" complement(3898. .4947) /gene="b116274" /note="ORF ID:b116274 putative dehydrogenase" /codon_start=1 /transl_table=11 /protein_id="BACS1539.1" /db_xref="GI:27354553"
CDS	/translations="MTGASDTVAIGADLRPPADLLRAGFTADQADETAAMLVWAN LRGDSHGLKIRIPRYVEMVEIINPAEPPOVSGKAVAVIADQARAGVMSLAAR KASBLADTHGKWCVRSISHAAGIYAHVOAQRIGIYMTASKPLMIYHSGSAG VSTNPIALAPFTGNDRPLLDMSGTASVALGKIMARQGPPIPGNAVDNGQATTD PREAKAVLPMAGPKGSGSLMI EYLA SVLTANPLI AVALVHGDPGSGGLVAVLDPSA FGDNFTLEVDRLGAIKSLPPAGADSVYLPGBERGFEEMERRSRAGITPLRAGTVSRL ALASKIYAVPEGVA"
gene	complement(4950. .5873) /gene="b116275" complement(4950. .5873) /gene="b116275" /note="ORF ID:b116275 hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BACS1540.1" /db_xref="GI:27354554"
CDS	/translations="MSAPPLSQPTREIPMOEKHTRFSDGLKLSAVLHVPDKRPGQRL PAFLVHGFSGKDSHAEIOARMLEQGYALRFDPCGSESEGERAQVCFDQVAD AKNALTFLEAREVDPRIGVYSGFAGVAVYAGVDRPACVSSCGMDGGERFR GOHPREMAKFTSLIEKRGKOKETGOSMISRDAVPIEBHRLKNSLPAYIMIPV ETAMSTRPADDEVGNIAFRPILFTHANTTITPTBOSIMFEKAGQPAELVLTGT SHPLAEDDAPRTKAIKGLDKFPSPPLGA"
gene	complement(5910. .6812) /gene="b116276" complement(5910. .6812) /gene="b116276" /note="ORF ID:b116276 unknown protein" /codon_start=1 /transl_table=11 /protein_id="BACS1541.1" /db_xref="GI:27354555"
CDS	/translations="NPIVVAHVGHKINTTIAEAVVPRAATEILGIQITYTBEGQWLS SNGRAELAVILRSHENATHTIGLEALTVDAVREASRVESACRVISKSPSLDCMAG VTFDPEGLRFEVHTPTDQIYGRYPTPHGVPNMDHLNLTSPDIATROQLERICG LRISBRMTDSSIMWYGGNROHILYGVVRGKGLHYHFEFEKFDYKLDLDRFD ROMLNGPRHRPGDNTVAYYTDSGAMIECGSMALLDDDGQFVNVNITNLERPQNR CMNVGTGPAPELMRQFHPFTAIT"
gene	7080. .7853 /gene="b116277" 7080. .7853 /gene="b116277" /note="GntR family ORF ID:b116277" /codon_start=1 /transl_table=11 /product="transcriptional regulatory protein" /protein_id="BACS1542.1" /db_xref="GI:27354556"
CDS	/translations="MILGSRGMAGRRMRBRVYRAGETRADVYLRMRADIIISCTLK PKAKRFEALRITAYVSGTSLREALSRVABGVYVAEORGVVAVPISIDLNDLTV RVLIERECVLAIKNGDAWEDITIGAFHRMDRLONRGSNTYLSSENGKLHGDHFS LVAAQSNLLLEIRKQLEFARHRYRMSOFRTKVRADVEHKMLTMSVVAADAKAE ELIEHRIETDNNYIKHAGHLFVTSDEBRPRTSRVAE"
gene	complement(7993. .8880)

CDS
/gene="11ve"
complement (7993.. 8880)
/gene="11ve"
/note="ORF_ID: b116278"
/codon_start=1
/transl_table=11
/product="branched-chain-amino-acid transaminase"
/protein_id="BAC51543.1"
/db_xref="GI:27354557"
/translation="MVPTNSISHSRTWTFPRGOMHEGNVPIMGRTHGLMLATIPD
GARAPEGVMPDLDRHCARNOAANVNGKLPVVDLTLGLVRBGIAFPVAELTIRP
MYWPONGIGGVLPDPEITDWCLESYPIEPVGSALTLSPPRPAACAPAVKAS
CLYNSSRALREASRGFONCLMDLNVTEFGNSNFMKADGVFPAPANGFLNG
ITRORVILELRGDGIVLEATIRYSDFVSADEISTGNFAKAPVIRIDRQLRGRF
YARAKLYWDFAHGKLA"

gene
complement (8909.. 9784)
/gene="b116279"
complement (8909.. 9784)
/gene="b116279"
/note="ORF_ID: b116279
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC51544.1"
/db_xref="GI:27354558"
/translation="MRVMTMRCEBIGIADLDAIADLLTRGEPTRRDYWDGLRRLS
VRETPDSFRRGYMLEHNRAPVAGALOLITHRNDGGSISTRCSMWVEBAPRSYAP
MLTKVAQRHNDVTVNINISPAETWYVIGGQFRRYCGSLFSPRLSPAPAMVEV
RQDAIDGLSEADVVLARHAAVGCCLICRASGRAPFLVDPMRVRLRLAMOM
IYCPDVAVVCAAGIQRFLRGRISIALDANGVRGLVGIYEBPIGRKFKPHCP
RLADLSDSELYYGP"
complement (9781.. 10938)
/gene="b116280"
complement (9781.. 10938)
/gene="b116280"
/note="ORF_ID: b116280
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAC51545.1"
/db_xref="GI:27354559"
/translation="MTVLIDVSDSEHLLSRAPNGARAMASPGNRACLDIGLNNM
SDAAMSTERQLFDLAAAGRLCYRLHFYMEATPREEMGRDYVRARYRGIIDLNR
RLDGVITGAPRAASLTDEPYMTFAIADARENTVSFVSCIAVGAVALHDGVT
RHKLPTGCFVFAOTKRHHPLMODVPTPRI PHARMVEDELDSDGYSVLMSAE
AGVCFYKQKNSLFYHFGHPEYETGSLSEYRDMKRFAGEHEVPTTIPRTYLNK
GAEBILIAFRQALSDROPFLPADPAPDQLTRDSNVHLPAOHLYRNWLOMVSOPA

Query Match 12.5%; Score 157.4; DB 1; Length 302650;
Best Local Similarity 49.4%; Pred. No. 6.4e-17;
Matches 556; Conservative 0; Mismatches 551; Indels 18; Gaps 5;

QY 49 TTCAAGCCGTTGGCGCTGGGATCGCGAGAGACCGGTTCCCGCGATGAGCGCTGAGA 108
DB 294417 TTCAATCCGCTGTCGCGGAGACTTCACCGCAACCCCTTACGATTAACGACGCGCTGGC 294358

QY 109 GAGGCAACCCCATTTTCTACGCGGATGAGGCGCGCTCTGGTCTCTACCCGATACAC 168
DB 294357 GCGATCGATCCGATCATGTGACGCGCTTGGCCAGTTTC--GTGCGCAGTGGCCAGCGC 294301

QY 169 GACGTGTCGGCGGTTTCCGCGACGAGCTTCCGCGTCACTCGAGAGAGTGGGATTCG 228
DB 294300 GATGTCAAGCTCTGTATGCGGACAGAGCGCTTCCGCGAAGATTTGTGAGAGCGGTCCAG 294241

QY 229 AGCGCGAGTACTGTCGCGCATTCGCCAGC--TCACGATATGAAAGTACGATTCG 285
DB 294240 CGCGGCTACAGCAAAAGATATGACGAGCGGCTGTCCGAGATGACCACTGATG 294181

QY 286 TTCGGCGTCGCGCGAGATACGCTGCGGTCGCGCAAGCTCTCTAACCCGCTGTTACG 345
DB 294180 CTCGAGCGCGATCCCGGACACACCGCGCTGCGGCTCTGTGTGTAAGGCCCTTACCC 294121

QY 346 TCACGCGCATGACTGCTGCGCGCGCAAAATACAGCGACCGTGCACGAGTGTCTCAT 405

DB 294120 GCGCGCGCGCTCGAGCATCGCGCGCGGATTCAGAGAGATGCTGATGAGGCCATCAT 294061

QY 406 GCTGCTCCGGAACAAGAGATTGACGCTTGTGGGGATTACCGGAGGAAATCCCATG 465

DB 294060 GCGGTATCGACCGCGGCAATGACCTGATTCAGAGATTTTCCTCCGCTGCGCTC 294001

QY 466 CGCGGATACAGCGCTCTGTTAAAGTTCCGCGCAGTGTGACGAAAGTCTCGTGCCTTC 525

DB 294000 ACCTATCTCGCATCTGCTGGAATCCCGAGATATCGGAGTCTT-----CTAC 293947

QY 526 GCGTCGCGACTGCGCGCGCGCTCGCGCTGGGTTTGTCGCCAGTGCATGAGAGAC 585

DB 293946 AAGAGCTGCGCGACGCGGCGGCGCTGCTGACCCCGGCGCTACCGCGAGAGATC 293887

QY 586 AAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 645

DB 293886 GCGAAGGCGCAACGCGCGGCAATGATGCGGAGATGATTTTCAGCAGCTGTTGAGCTC 293827

QY 646 CGCGCGAGGAACCGCTCGAAATGACGCTTGAAGTGTGCTTCAGCGCGGCGGCGGAC 705

DB 293826 CGCGCGCGCAATTC--CGCGAGACTTACACAGCTGTGTGAGGCGGAGAGAC 293770

QY 706 GCGACAGGCTGAGCAGAGAGAGCTGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 765

DB 293769 GCGAACAAGCTACCAACAGAGAGCTGACGCGCAACATCATCTGTTGCTGCGCGCGG 293710

QY 766 ACCGATACACAGATCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 825

DB 293709 CACGAGACCAACGCTCAATCTGATGCGCAACGCGCTCTTGCGCTGCTGCTGCTGCTG 293650

QY 826 GCGCTGAGCTGTGTAAGGCGCGGCGGCTCATGAGGAACGCGCTGATGAGGTGCTC 885

DB 293649 CAGCTGCGCTGCTCAGAGCGCGGCGGCGGCTGATGCTGCTGCTGCTGCTGCTG 293590

QY 886 CGCTTCGACCAATATCTCAGATGAGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTG 945

DB 293589 CGTACGATCTCGGTGAGATGACCGGCGGCGCTGACGCTGAGAGATGACGATCTC 293530

QY 946 TCGCGGCGATGATCAAGAAAGGAGATGCTTCTTCTGATTCGAGCGGCTGAGA 1005

DB 293529 GCGCGCAGAAATCCCAAGGAGGAGCCGCTGTGCTGCTGCTGCTGCTGCTGCTG 293470

QY 1006 GATGGAGCTGATTTCTCC--AGCGCAGAGCTGTTTGAATGCGCAGCGGACAGCGGCG 1062

DB 293469 GATCGCGGCTGCTTATCCGATGCGCGCGGCTGCGAGCTGACACCTGCTCAAG 293410

QY 1063 AGCTGCGGTACGCTGAGAGCGCCCATGCTGCCCCGCGGCTGCTGCTGCTGCTG 1122

DB 293409 CGCGTGTGCTTCCGCGGCGGCGATCATTTGCTGCGGCGGCGCAATTGCGCGATGAG 293350

QY 1123 GCGGAGATCGCGGTGGGACCATCTTCCGTAGGTTCCCGGAGATG 1167

DB 293349 GCGGAGATCGCATGCGCACGCTGTGCGCGGCTGCGCGATG 293305

RESULT 12
SERCP450A 3078 bp DNA linear BCT 26-APR-1993
LOCUS Saccharopolyspora erythraea ORF 1 gene, partial cds; cytochrome
DEFINITION P-450 gene, complete cds; ORF 2 gene, partial cds.
ACCESSION M81110
VERSION M81110.1 GI:152682
KEYWORDS cytochrome P450.
SOURCE Saccharopolyspora erythraea
ORGANISM Saccharopolyspora erythraea
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Pseudonocardiales; Pseudonocardiaceae; Saccharopolyspora.
REFERENCE 1 (bases 1 to 3078)
Andersen, J.F. and Hutchinson, C.R.
Characterization of Saccharopolyspora erythraea cytochrome P-450
genes and enzymes, including 6-deoxyerythronolide B hydroxylase
J. Bacteriol. 174 (3), 725-735 (1992)

MEDLINE	92121109
PUBMED	1732208
COMMENT	Original source text: <i>Saccharopolyspora erythraea</i> (library: NRRL 2338) DNA.
FEATURES	location/Qualifiers
SOURCE	1..3078 location: "Saccharopolyspora erythraea"

CDS
<1..859
/note="ORF homologous to Escherichia coli and Salmonella
typhimurium dnaE gene; putative"
/codon_start=2
/transl_table=1
/protein_id="AA026482.1"
/db_xref="GI:152683"
/translation="AGGCGGHRQAQAPAGARFVRPYRRKRRRRRRVVVAAGAPVHAHG
RVAASALDLRAGDAGPLRLGPPSGRGAARAAVYGHRRARRRRRAAGCGSPQDERR
DLGGPAPHOGERPVGDRDPCGPRGSGAVLPEVLGIDVRLPARGHRRDGEHPDRR
ARGHIOHLRRRGAGAGLGGRRPQGNORVNDQSGEGVAGRRRAOHLAAGLGGH
PGARPAAGAGHRPPRAVQRLRLVDRERAAARAEAAARGVLRGCAVSRRAASPLEAA
RRPLRLRGS"
949..2166

```

/codon_start=1
/transl_table=11
/product="cytochrome P-450"
/protein_id="AA25483.1"
/db_xref="GI:152684"
/translation="MTTGEVDPDLAFDDAPADQRRHNRVARRKEEVPQRIKTNGLDAM
LITREDVQKALLDPIRAIKDFGRTOQII EKLALAEKPPGSPGLGPMNTDTPDHIT
RLKRVVFAFARVEGRVAPRIEQITDILDRLAGRSVDIIDEAPFLPTIVSIALEE
GVEDSRDVFSTWTVLVDSQPEAKQASVAMVEYLTLEIAKRTGEGDQDLSLAE
AVEEDRLSEGLIMVFLVLAHETTIVNLGVNLISLGNPOLAALRNDPDLAL
AIEELTRVESPANGTFRHTAVAPFGVVVPEGLVVALGANPQDSEDRDDRD
ITRETTGVARGHGIFCVGALALRLAQIAYGRLLERFPLRLAAASDDLKRMFSVL
MAGLEKLIVRPGA"
complement(2365..>3078)

```

```

CDS
Complement(2355..35078)
/note="ORF 2"
/codon_start=1
/translation_table=1
/protein_id="AA26484.1"
/db_xref="GI:152685"
/translation="LIVATGSGTLFVLSOYLIVRGSHAVVAGMAALPIAAGSVVSGA
LGAAPARICGACITGCAATAGCGCTAAGGDSGCPHAPGLIGCTGCPAGPA
ATATLCAVPPDRAGMSGALADTHOQGIAGVAVLGSLSTAIRAPLIPVPEPDAST
VLTSTPADERTSALADARLARLFTQAGSATMTAGLACALAGAAVMILSTGRASA
VLAIVASVATARRINPAGGT"
BASE COUNT      463 a      1060 c      1137 g      418 t
ORIGIN

```

Db	1209	GAACCTGGGGCCGGCAATGCTCAACACCGAACCGCCGGACCAACCCGGTTGGCCAAAGCT	1268
Qy	327	CGTCACCCCGTCGTTTACGTACAGCGCCATCGACCTCTGCGCGCCGAATATACAGCCAC	386
Db	1269	GGTGTTCMAAGCGCTTCAACCCCGCGCGGGATCGAGGGCTTCGCGCCGCGAGATCGAGCAGAT	1328
Qy	387	CGTCGACCACTCTGTGTGATCTCTGCTCCGGACCAAGAGATTGTGACGTTGTGGGGATT	446
Db	1329	CACCGAGACCTCTGTACACCGCCCTGGGGGGCCGGTCCGAAGTGGATCTGATCGACGAGTT	1388
Qy	447	CGCGGAGGGAATCCGATATGCGCGCGATCACCGCTCTTTAAGTTTCCGGCCGAGTGTGA	506
Db	1389	CGCCTTCCCGCTGCCATCAACGGATGATCTGGAGACTATGGCGGTGAGAGAACTCGCGCG	1448
Qy	507	CGAGAGTTCTCGTCCTTCCGCTTCGCGGACTGCGCGCCCTCGGCGCTGGGATTGTGTCC	566
Db	1449	CGAGCACTTCCGG-----TGTGTGACCAACGTCCTGTATCGACGGATTCGACGCCGA	1499
Qy	567	CCAGGTGGAATGAGAGAACCAAGACCCTGTGTGCGTCCGTACCGAGGGGGTCCGCGTGT	626
Db	1500	GGCGCAGAGCCAGGCGCTGGATGGAGATGTGATCTACCGAGCTGATTCGCAAGAA	1559
Qy	627	CCATGACGTCTCGATATAGCGGGCGCAGAGAACCGCTGAAATATGACCTTTTGACGATCT	686
Db	1560	GGCGACCGGACCCGGTATC-----GACCTGCTGACCGCGCT	1595
Qy	687	GCTTCAGGCGCAGGCGCAGCGCAGCAGAGCTGAGCAGAGAGACTGTGTGCGCTGTGG	746
Db	1596	GCTGAGAGCGGTTCGAGAGACGGCGCACCGCTCAGGAGAGGGAGAGCTGATTCGCAATGGT	1655
Qy	747	TGCGATTATCGCTGTCTGTGGCACCGATACCAAGATCTACTTATGCGCTTCCGTGTCTCA	806
Db	1656	CTGTCTCTGTGTGTGCGCGGCGCAGAGAGACGATCTCACTGATCGGCAACTGTGCTGTCTC	1715
Qy	807	CTTGTCTGCGGTCCGCGAGGCGCTCGAGCTGTGTGAAGCGCGAGCCGGGCTCATGAGAGAA	866
Db	1716	GCTGTGGGGCAACCCCGACAGCTCTCCCGGCTGTGCGCAAGAACCCGTCTGCTGCCCG	1775
Qy	867	CGCGCTCGATGATGATGTCTCGCTTCGACCAATATCTCAGAATATGAGAACTGTGCGTTTCG	926
Db	1776	CGCGATGAGAGAGACCTTGGCTCTACGAGACCCGGTGTGGCAAGCGCAACGTTTCCGGCACAC	1835
Qy	927	CAGGCAAGAACCTGAGATTACTGTGCGGGCATGATTCAGAGAAAGGGAGATGTTCTTCTCT	986
Db	1836	CGCGGAGGCGGTTCGCGATTCGCGCGACGATGTATCCCGAGAGAGAGCTCTCTGTGGGTCG	1895
Qy	987	GATCCGAGGCGCCCTGAGATGAGGAACTGATTTCTCAGGCGCAGACGTTTGTATGTGCG	1046
Db	1896	GCTTGGGGGCGAGCAACCGGAGCGGTGAGCCCTTCGAGAGACCCCGACCGCTTTCGATAC	1955
Qy	1047	ACGGGACACGGGGCGAGACTTCGCGTACGGTATGAGGCCCCCATGTTCTGCCCGGGATGT	1106
Db	1956	CCGGGAGAACACCGGGGACGTTCCGCTTTCGGCCACGGGATCACTTCTGCGTGGGGCGCGC	2015
Qy	1107	CTTTGCTGCGCTTCGAGCGGAGATGCGCTGTGGGCAACCATTTCCGATGATTTCCCGAGAT	1166
Db	2016	CTTGGCCCGGCTGGAGGCCAGATTCGCGCTGTGGAGGCTCTGTGAGCGCTTCCGGACCT	2075
Qy	1167	GAAGCTGAAGAAACTCCG	1186
Db	2076	GGCGATGGCGCGTGTGCGCG	2095

[illegible]

REFERENCE 1
 Micromonosporineae; Micromonosporaceae; Micromonospora.
 Farnet, C.M., Staffa, A. and Yang, X.
 Genes and proteins for the biosynthesis of rosamycin
 JOURNAL
 Patent: WO 03010193-A 7 06-FEB-2003;
 Ecopia Biosciences Inc. (CA)
 FEATURES
 source
 1..1233
 /organism="Micromonospora carbonacea"
 /mol_type="genomic DNA"
 /db_xref="taxon:47853"
 BASE COUNT 225 a 452 c 377 g 179 t
 ORIGIN
 Query Match 11.9%; Score 149.4; DB 6; Length 1233;
 Best Local Similarity 49.7%; Pred. No. 3.6e-15;
 Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;

Qy 288 CGGGCTGCGCGCGAGATCAACGCTCGGAGCTCGTCAACCCGCTGTTACGTC 347
 Db 261 CGGCATCGACCCCGCGCGACGCGCGCTGCGCAAGCTGTGAGCCAGCGGCTTACCCC 320

Qy 348 ACGGCCATCGACCTGCTGCGCGCGCAAAATACAGCGCACCGTTCGACAGCTGCTGATGC 407
 Db 321 CGCGCGATCGCGCGAGATGAGCTCGGATCGGCGAGATCACCGCGGAGCTGCTGACCA 380

Qy 408 TCGCTCCGAGCAAGAGAGTTGACGTTGTGCGGATTAACGCGGAGGAATCCGATGCG 467
 Db 381 GGTACGCCACACGACCGGATGACATCGCCAGCACTCGGTACCGCGCTGCGTAC 440

Qy 468 CGCGATGACGCTGTTGAAAGTTCCGCGCGAGTGTGACGAGAAATTCGCTGCTGCG 527
 Db 441 GGTATCGCCGAGCTGCTCGGATTCACCAACAGATACAGAGATTCCGCGAGTGGGT 500

Qy 528 CTGCGGATCGCGCGCGCTGCGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 587
 Db 501 GACATCTATCTTCAGCAACGAAAGGCTGAGATCCCAACCTCCGAGCACTTACCGCA 560

Qy 588 GACCTGTGCGGCTCGTCAACGAGGCGCTCGGCTGCTCAATGACGCTCGATGAGCG 647
 Db 561 GACGCTGCGCGCGCGATGAGAGTGTGCGAATTCCTGTAAGCCAGATCGCCACAA 620

Qy 648 GCGCAGAAACCGCTCGAAATGACGCTTGAAGTGTGCTTCAAGCGCGAGCGCGACG 707
 Db 621 GCGCGCGACCG---AAGAGACCTGATGACGCGCTGTGCGGAGGATGAGCG 677

Qy 708 CAGCAGCTGAGCAGAAAGAGCTGTGCGCTGTGAGTATGCTGCTGCTGCTGCTGCTG 767
 Db 678 GCGCAAGCTGACCGAGAGAGTGTCAACATCGTGTGCTGCTGCTGCTGCTGCTGCTG 737

Qy 768 CGATACACGATCTTATGCGGCTGTGCTGCTCAACCTGCTGCGGCTGCGCGAGCG 827
 Db 738 CATCTCCAGCGCGACGCTGCTGAGCAACCTGTTCTGTGTGAGAGACACCGCAAGC 797

Qy 828 GCTCGAGTGTGAGAGCGCGCGCGCTCATGAGAAACGCTGCTGATGAGTGTGCTGCG 887
 Db 798 AAGCGCGCGGCTCGCGCGCGACCGAGCTGTGCGCGCGGCTGATGAGAGAGCTGCG 857

Qy 888 CTTCGACAAATCTCTCAAGATGAGAACTGTGCTTTCGCGAGGACGAGACTGAGTACTG 947
 Db 858 CT---ACGCGTCCCGTTCACTGATCTTCGAGATCTTGAACGAGGACACCGATCTCT 914

Qy 948 CGGGGCGATCGATCAAGAAAGGAGAGATGCTTTCTCGGATCCGAGCGCGCTGAGGA 1007
 Db 915 CGGCGACCCGAGCGAGAGGCGAGATGTGATCGCTGAGATCGCTTCGCGCAACCGGCA 974

Qy 1008 TGGGCTGATTTCTTCAGAGCGAGAGTGTGAGTGTGAGCGGAGCAGCGCGAGAGCT 1067
 Db 975 CACCGAGGTGTTCAAGAGACCGGAGCACTTGAACATCGAGCGGATGAGAACACACTT 1034

Qy 1068 CGCGTACGTTAGAGCGCGCGATGCTGCGCGGCTGTGCTGCTGCTGCTGAGGCGGA 1127
 Db 1035 GCGCTTTCGCGCGCGATCGACCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1094

Qy 1128 GATCGCGTGGGACACATCTTCCGTAGGTTCCCGAGAT 1166
 Db 1095 GGTCTTCTCAACGAGAGCTCGACGATTACCGAGAT 1133

RESULT 14
 AX697977/c 60196 bp DNA linear PAT 02-APR-2003
 LOCUS
 DEFINITION Sequence 1 from Patent W003010193.
 ACCESSION AX697977
 VERSION AX697977.1 GI:29499034
 KEYWORDS
 SOURCE
 ORGANISM
 Micromonospora carbonacea
 Micromonospora carbonacea
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE 1
 Farnet, C.M., Staffa, A. and Yang, X.
 Genes and proteins for the biosynthesis of rosamycin
 JOURNAL
 Patent: WO 03010193-A 1 06-FEB-2003;
 Ecopia Biosciences Inc. (CA)
 FEATURES
 source
 1..60196
 /organism="Micromonospora carbonacea"
 /mol_type="genomic DNA"
 /db_xref="taxon:47853"
 BASE COUNT 6934 a 20042 c 23822 g 9398 t
 ORIGIN
 Query Match 11.9%; Score 149.4; DB 6; Length 60196;
 Best Local Similarity 49.7%; Pred. No. 1.8e-15;
 Matches 437; Conservative 0; Mismatches 436; Indels 6; Gaps 2;

Qy 288 CGGGCTGCGCGCGAGATCAACGCTCGGAGCTCGTCAACCCGCTGTTACGTC 347
 Db 3601 CGGCATCGACCCCGCGCGACGCGCGCTGCGCAAGCTGTGAGCCAGCGGCTTACCCC 3542

Qy 348 ACGGCCATCGACCTGCTGCGCGCGCAAAATACAGCGCACCGTTCGACAGCTGCTGATGC 407
 Db 3541 CGCGCGATCGCGCGCGAGATGAGCTCGGATCGGCGAGATCACCGCGAGCTGCTGACCA 3482

Qy 408 TCGCTCCGAGCAAGAGAGTTGACGTTGTGCGGATTAACGCGGAGGAATCCGATGCG 467
 Db 3481 GGTACGAGACAGAGCGCGAGTGAACATCGCAGAGACTCGGTACCGCGCTGCTGAC 3422

Qy 468 GCGCATGAGCTGCTGTTGAAAGTTCCGCGCGAGTGTGACGAGAAATTCGCTGCTGCG 527
 Db 3421 GGTATCGCCGAGCTGCTCGGATTCACCAACAGATACAGAGAAATTCGCGAGTGGGT 3362

Qy 528 CTGCGGATCGCGCGCGCTGCGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 587
 Db 501 GACATCTATCTTCAGCAACGAAAGGCTGAGATTCCAACCTCCGAGCACTTACCGCA 3302

Qy 588 GACCTGTGCGGCTCGTCAACGAGGCGCTGCGCTGCTCATGACGCTGCTGATGAGCG 647
 Db 5301 GACGCTGCGCGCGCGCGATGAGAGTGTGCGAATTCCTGTAAGCCAGATGCGCCACAA 3242

Qy 648 GCGCAGAAACCGCTCGAAATGACGCTTGAAGTGTGAGATGCTTCAAGCGCGAGCGGCG 707
 Db 3241 GCGCGCGACCG---AAGAGACCTGATGAGCGCGCTGTGCTGCGGAGGATGAGCA 3185

Qy 708 CAGCAGCTGAGCAGAAAGAGCTGTGCGCGCTGTGAGTATGCTGCTGCTGCTGCTGCTG 767
 Db 3184 GCGCAGCTGAGCAGAGAGAGTGTCAACATGTCGCTGCTGCTGCTGCTGCTGCTGCTG 3125

Qy 768 CGATACACGATCTTATGCGGCTGTGCTGCTCAACCTGCTGCGGCTGCGCGAGCG 827
 Db 3124 CATCTCAGCGCGCGAGCTGCTCAGACACTGTTCTGTGTGTTGAGAGAGCAGCCGAGCG 3065

Qy 828 GCTGAGCTGTGAGAGCGCGCGATGCTGCGCGGCTCATGAGAACGCGCTGATGAGTGTGCG 887
 Db 3064 AAGGCGCGGCTGCGCGCGAGCGCGAGCTGCTGCTGCGCGCTGATGAGAGAGCGCTGCG 3005

QY 888 CTTCGCAATATCTCAGATAGATACTGTGCTTTCCGAGGACGAGACTTGAGTACTG 947
DB 3004 CT---ACCGGTCCTCCCTTCACTGATCTCCGAGTCTGAACGAGACACGACATCTCT 2948
QY 948 CGGGGCGATCGATCAAGAAAGGAGATGGTCTTTCTCTGATCCGAGCGCCCTGAGAGA 1007
DB 2947 CGGGACCCCATCGGAGAGAGGCGGAGATGATGCTGATGCTGATGCTGCTCCGCAACCGCA 2888
QY 1008 TGGGACTTATTTCTCCAGGCGAGAGCTGTTTGAATGCGACGGAGACGGGCGGAGCTT 1067
DB 2887 CACCGAGTGTTCACGAGACCGGACACCTTCGACATCCAGCGGAGTCAACAGACACT 2828
QY 1068 CCGCTACGATAGAGGCGCCCATGCTGTCCCGGGGCTGCTTCTGCTGCTCGAGCGGA 1127
DB 2827 GCGCTTGGCGCACGCGGATCCACCATCTGCTGGCGGGCTTCGCGCAGGCTGAGGCGAA 2768
QY 1128 GATCGCGTGGGACCATCTTCCGATGCTTCCCGGAGAT 1166
DB 2767 GGTCTTCTCAACGACGCTGACCATCTTCCGAGCTT 2729

RESULT 15
SAU421825/c 66808 bp DNA linear BCT 16-MAR-2002
LOCUS DEFINITION Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
ORF1, stia gene, stib gene, stic gene, stid gene, stih gene, stlf
gene, stlg gene, stih gene, stih gene, stih gene, stih gene and
ORF9.
ACCESSION AJ421825
VERSION 1
KEYWORDS acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;
ORF7; ORF8; ORF9; polynucleotide adenylyltransferase; potassium
channel beta chain; ribosome binding factor; stia gene; Stia
protein; stib gene; stib protein; stic gene; stic protein; stid
gene; stid protein; stih gene; Stih protein; stih gene; stih
protein; stih gene; Stih protein; stih gene; stih protein; stid
gene; Stid protein; stik gene; stih gene.
SOURCE Stigmatella aurantiaca
ORGANISM Stigmatella aurantiaca
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Cystobacteriales; Cystobacteraceae; Stigmatella.
REFERENCE 1
AUTHORS Gaiatzis,N., Sikowski,B., Kunze,B., Nordsiek,G., Blocker,H.,
Hofle,G. and Muller,R.
TITLE The biosynthesis of the aromatic myxobacterial electron transport
inhibitor stigmatellin is directed by a novel type of modular
polyketide synthase
JOURNAL Online Publication
REMARK J. Biol. Chem., 10.1074/jbc.M11738200
REFERENCE 2 (bases 1 to 66808)
AUTHORS Muller,R.
TITLE Direct Substitution
JOURNAL Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
Braunschweig, GERMANY
FEATURES
source location/Qualifiers
1..66808
/organism="Stigmatella aurantiaca"
/mol_type="genomic DNA"
/strain="Sg a15"
/db_xref="taxon:41"
complement(4978..6231)
/note="ORF8"
/codon_start=1
/transl_table=11
/product="cytochrome P450 monooxygenase"
/protein_id="CAD19077.1"
/db_xref="GI:19572310"

CDS
/transl_table=11
/protein_id="CAD19083.1"
/db_xref="GI:19572316"
/translation="MRPRIRIGAPGAVPILAMARVAPGAVKQQLAROPYLAHVA
AMPDVHVAEGVAVGVFATERHIVVAGLGDGCGVSAVRFDGPAMSIGRVELTRILA

PGDPATDVLTRLIQESDGERLTETELLHOCFILNAGHETTNLIGNALLEAREPD
ERARLARAPALIPTRAVEVLIYBESNOIGNRVADPEDIIGVAVPTEPLTTCIGAN
KDPAREFDEPHEDVGRONRHARFGGHTCAKMLNARIETARIALAPLARPPVALT
APVAVARRARFRFGTAMPTRLGSLK"
6369..7730
/note="ORF7"
/codon_start=1
/transl_table=11
/product="protein kinase"
/protein_id="CAD19078.1"
/db_xref="GI:19572311"
/translation="MEELISBALPDGSHVCPMRLEGAGYGTGAVVYAHVGRICPR
VALKLARYPNDLREAREALLERYRHAIVLETHMGCLHMDVKGENVLYSPGHSIMD
VRLYOMKREPLRTAREVLRVLAOVARALEATHMGCLHMDVKGENVLYSPGHSIMD
FGCTWEGAPLPTGILPRTKARSPALFPHNHRBAADHATATPADVYALGVT
AYRLCTGIYPPATPEISGVSDERTQVKKPQSVQRLVPALBQLILRMIDNKER
GRAGELAEEMELAEARAGADEPMPNPHSSVPSBITRPSVQREMGGWPAVLVA
VGLILMGSGSLNTRVAVAPASRMTEGGVADAVALPVSARAREPESGLNDML
KEPLPGQRRPCCPRQISIQGGCWLVEVPQSSPCGESYEWKAGACYSVAAPRRPTS
DPO"
complement(7849..8577)
/note="ORF6"
/codon_start=1
/transl_table=11
/product="methyl transferase"
/protein_id="CAD19079.1"
/db_xref="GI:19572312"
/translation="MSONIYDDPSFESYQLASLKLGAAPEMPALRALIPVKDL
RVVDLGCQGFRCWARNDAAHVAGLDLSKMLTRAETSGDAITYQOADLETLV
PEGADLAFSSLAHYVELEPLATVIRGLTPCGWFVFSIEHPFMASROPEWMTGP
EGHRTMPVDSYOLEGKRTTHWLAEEVVKYHRTLTGTTNALLEGQFTLHIBWGPDA
OLSAPELAEBSRDRPMFLVSAQR"
complement(8622..8879)
/note="ORF5"
/codon_start=1
/transl_table=11
/product="acyl-CoA binding protein"
/protein_id="CAD19080.1"
/db_xref="GI:19572313"
/translation="MALEDPRSAQERVKALKTRPSNDRLLELYSLFKQATEGVQK
RPGIIVDERAKYDAMVGRKLGRESAMQYVALVERLLRG"
8861..10070
/note="ORF4"
/codon_start=1
/transl_table=11
/product="potassium channel beta chain"
/protein_id="CAD19081.1"
/db_xref="GI:19572314"
/translation="WVYLCSRSQLRWLNGSAHRVRFAMGHRPRMRHRLGRS
GLMVEISYGNWLTGSOVEBAANAVCYRAALDEGITTPTADVYAGTRAEVIGRAL
OGOREGIEITRTYWPFTGPNDRGLSRKHLBSIHSLRKTETVYDLYQAHRYDT
ETPLEETQAFSDVVRQGRKAHYIGSEHTAQLRGVLSQMGFQLVSSQFQISMVM
RVILEAVVPASQELGQGVVSPIAQGVLTGRYRQGRPPGSRATIDKGASFIRK
LNDVLTIVQOOLTPVAGBAGLSLAQLAAMVLYONPNVSSAIIIGASRPQOVENVKAC
VKLSEAKRHIDTLTGPIVERNPKTKLENSRKRKP"
10428..10862
/note="ORF3"
/codon_start=1
/transl_table=11
/product="ribosome binding factor"
/protein_id="CAD19082.1"
/db_xref="GI:19572315"
/translation="WTSFKQRRRARPRGCTSSLSPPAQSLPAHHLAVOSTLPQEV
RLFRGLSDPLEGVSVTSFELTPGKNARKGFTLPPEYABGSGAPITAAVVRASGFI
RAQLALGNTLRKVPPLRFVCVAPPGFSIPEDGSAEGBGS"
10859..12016
/note="ORF2"
/codon_start=1
/transl_table=11
/protein_id="CAD19083.1"
/db_xref="GI:19572316"
/translation="MRPRIRIGAPGAVPILAMARVAPGAVKQQLAROPYLAHVA
AMPDVHVAEGVAVGVFATERHIVVAGLGDGCGVSAVRFDGPAMSIGRVELTRILA

CDS

```
AFARAVPVGADVHRGRGLSLPGLDIGRLSTHRLQREWERLVPKHLGLTGGNHLEL
DRDEBAGMLTLHTSGRGVGAIAAHNLVRAQVSGEGLPGRLTSTSEGAACLDTEL
ACOFBARBDTLLOBAVEVLAGAIGQGVLPVAVHNVHVAASHFERTLHVRKA
VGVAGQOGLRGSWGTAYLVAGGEPOACSCSHGGRVTRERARERLRPALAH
ALRYVPDEGRVTALVEAPTAAYRDTITVELEDELVTPLRLTPIAVLKG
complement (12082, .13974)
/feature="ORF1"
/codon_start=1
/transl_table=11
/product="cellulase"
/protein_id="CAD19084.1"
/db_xref="GI:19572317"
```

gene

CDS

```
14345..21466
/gene="stia"
13345..21466
/gene="stia"
/function="PKS"
/codon_start=1
/transl_table=11
/product="Stia protein"
/protein_id="CAD19085.1"
/db_xref="GI:19572318"
```

```
/translation="MTOLASRLGVGSESIDPREPSRYGIDSRDKSLIAELALLKR
PLSPTLANTYPTVEALSHRLAGEADVSGEERADTGVQDEPIALIGAGFPKAPDV
ESFSLRDGVDALTEVPERMDLAPFYSVGLKSGKNTMGFLDQDQDPDLFG
ISPREAIQMDPOORLALSWELAESAGVAVDSLMDSPGVMGAMGSDALTEGAL
GLIGHPTAGDSDSIITARVSYFLGLRGLPVTNACSSALVAVALACOSLRNGEAM
ALAGVNLISPRESSVMSQCALSPDRCAFDARANGYRGEGBGVVITPLSRAT
ANGYPLVCVIRGSAVNDGFSNGLTAPNRPQOEVLRAVAVAGVSPKRVYVHTG
GTVLDDPIEANAASVILGVNPAERPLILGSKYNIIGHLEAASGAGMLKVL5IOHR
AIPMLHFEQPNPPIREFEELRLVPKSLHPNGBESALAGISFGCTGMLVLEA
PSPGSAHAPLAETREBELSLARQAPAVSVGSOVLADCSFVARCEGSRRLA
LIARPAELAVOAGFLAGEKRGILREBTVSGSQKPGVSGKPGVSGKPGVSGKPG
TEPVRGALIERCDRALRGISGKSHLIELEAGASRLKECSGLQRTAIRAIOVGLAAL
KSWGEPAGVGHSTIGEVAHVAGILNELEDAVTHLSRLMGLSQQMIIVDLS
AEBANRLREWERHVVAAACNAPASTVSGAREVIEBELAEETAAVYRNRPLTDS
HSPMLDPMEGLAEVLACISPRVAVPWSVTVGAFLEAEECTAAVYRNRPLTDS
OALTLEFODRLPLELSPHPLIARSVOCCKEGROGVILASMRGENEGNCESEL
GRLPTLGTSMQALYPAEAKRVELPMGLGTGSLDAELRGARLLPLSAHTAELEDA
ARALQASIAKHPVDLIDLCYSASVGSVYHARLAPYDTSKALVEHLDLTKGSRP
GRIGAVCATGEKRVFVFPQGSQMGQKQLRQEFREAIIECDALRSCGWS
LIDALQADEPHSRIGSIEVIOPTLFAIVALALMRSGVVDVAVGSGEATACV
AGALRLTDAARVTCERSQLRRVAGQGGALVELSEBAREMLSKRDERVIGASNP
RSTVSGDAIALEBILRELEAOGTCGRVAVDASHPOVPLKODLRVISTAPRA
ATVPLSTVGTGEPTEIGARMDSYMMRNLREPLFGOALERLAGDGHGVFLSPHPL
LPSIDQTLRHSHKVVPLSLAREEPRAVILLESIGTLFTGGGRVDFRGLHPLGHRV
RLPLHPWORQRFWLDMDVAAAGASVGAHMPGGS EHPDLGHRPLSLAHAPCHVE
NKLRSVASYLDDHRVGGTVPLGTSYIEMALAAAKSGKTKVSLRVAERKXSLA
ETASTVTOISLIOEBAGHSFRLHGRSTPDEPWTLAGGTVLAVSGEDRPPADPYLA
AVEYTRLOQKEVASADAYEAMWAGNEVGPRLGIBOMVRAELVALRVADEPLEOS
LPEYRHPAALDIDIGLOALIVLAHSGCGIALPRLSIEBAHSGGAGMCCVRLNPV
REASFESEVFLIDESGORLAELOGARQIIGESRPHDGRULVDMHLEVERQAPVP
VPTABSSQGVNLFADGTGVAEGLAQIIAAGBPCIVLSPRAAELEDEGFPVDE
QPEBVRKLFAGLSTSTIPLRGIVHLSLIDAPSPEBGVDLSPRAAELEDEGFPVDE
GLNAAELAQKPLMLVTRGVQVAGSAGAGAVAPSPMGQGVVAVHAELRCRVD
LAPVAPGEBOTLLOELLAPAGDOVALRGREHVRPLRGTCDGVAVASRDLRGAID
RSDGCVLTGGLGGLTIAKVLVARGABHVLVNGRBAVLAALALOMGAGAECL
VMKADVANTSEVEBELVGTIQRKMPPLRGITFPAAGLIDGTIVLQTLQALRYMAKVE
GAMVHRAITLASLDFVLFSSASLVVIGPGGANYAAGNAFLDALAGLRKARGACQS
IOMGWSBEGILAABQDNNGERLALRGMSLREBELALELISGEBPLREARIGVL
RLDARRMCKSYPAIRSSPLFVYLAGEBNAKVLERSGILRALNBEPARTRAALEF
```

LREBAQVLRIPPARIDLSRLQTLGNDLSLLELRNRLEANLELTPATVLMVPTL
SALAPFLAREIGIOLNVEBPQAVESPSVSEEDGAAMLKMLEOFOLISNBDLGLMLS
DDKXDEBVS"

gene

/gene="stia"

Query Match 11.8%; Score 149.2; DB 1; Length 66808;
Best Local Similarity 49.2%; Pred. No. 1.9e-15;
Matches 450; Conservative 0; Mismatches 458; Indels 6; Gaps 2;

QY	298	CCGAGAGATCAACGCTCGGCTCCGCAAGCTCGTCAACCGCTGTTTACGTACCGGCCATC	357
DB	5928	CCGCGCGCTGCACACGCGGCTCGCCGCACTGATCTCGGCGCGCTGACCCCGCGCGCTTG	5869
QY	358	GACCTGCGCGCGCGCAATACAGGACACCGTCCGACGCTGCTGATGCTCGCTCGCGCA	417
DB	5868	TCGACATGAGCCAGGAGGCTTCCGACCTTGATGACCGGCTGCTGACCGGCTGCGCTTG	5809
QY	418	CAAGAGAGATTGACGCTGTCGCGGATTAACGCGAGGGAATCCCGATGCGCGCATCAGC	477
DB	5808	AAGGCGCGCGGACACCTGATCGAGAGCTTCCGCGCGCGCATCCCATCGAGGTAATTGGC	5749
QY	478	GCTCTGTTGAAGTTCCGCGCGAGTGTGACGAGAACTTCCGCTGCTGCGCTCGCGCACT	537
DB	5748	AACCTGCTGACATCCCGACGAGAGAGCGGCGCGCGCTGCGCGCTGCTGCTGCGCATC	5689
QY	538	CGCGCGCGCTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	597
DB	5688	CTCGCGCGCTGCAACCCCGCTGACGCGCGAGCGAGGCTGCGGGAATGAGCGGTA	5629
QY	598	GCCTCCGTCACGAGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	657
DB	5628	ACAGAGTTCTGACATCTCGCATCTCGATCTCGATCGCGAGCGCGAGCGCGCGCGGAGC	5569
QY	658	CCGCTCGAAATGACGCTTGTACGATGCTGCTTACGCGCGGAGCGCGGACGCGAGCGCTG	717
DB	5568	CC---CGGACGAGCGTGTCTACCGCGCTCATTCAGCGCGGAGATCGAGCGGGAACGATTG	5512
QY	718	AGCAGAGAGCGTGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	777
DB	5511	ACCGAGAGAGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	5452
QY	778	ATCTACCTTATCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	837
DB	5451	ACGACCTCATTCGCGCAACCGCTGGAATGCTGCGCGCTTCCGGAAGAGCGCGCTCGG	5392
QY	838	GTAAGCGCGAGCCCGGCTCATGAGAACCGCGCTCGATGAGGTGCTCGCTTGACAAT	897
DB	5391	CTGCTGCGAGCCCGCGCTGATCCCGCGCAGTGGAGGATCTTCCGTTACGAATCC	5332
QY	898	ATCCTCAGAAATGGAATCTGTCGCTTTCGCGAGGACGCTGGAAGTACTGGGGGCGATCG	957
DB	5331	TCCAACGAGCTCGGCAACCGCGGAGTGAC---GAGAGACAGAGATTGGCGGCTTACGCG	5275
QY	958	ATCAAGAAAGGAGATGCTTCTTCTGATCCGAGCGCGCTGAGAGATGGGACTGTA	1017
DB	5274	GTCGCGACCGGCACTTCTCAGCTGCAATCGGCGCGGCAACCGGACCGCGCGGCGC	5215
QY	1018	TTCTTCAGGCAAGAGCTTGTGATGTGACGAGGACACGCGCGGAGAGCTTGGCTGATCGCT	1077
DB	5214	TTTCAGGACCCAGAGCACTTGCAGAGTGGGGGCGGACCGCAACCGGCACTGCGCTTCCG	5155
QY	1078	AGAGGCGCCCATGCTGTCGCGCGGAGTGCCTTGTGCTGCTGAGAGGCGGAGATGCGCGTG	1137
DB	5154	GAGCGGCGGCAATACGTCGCGGAGATGAACCTTGGCGGATAGAGCGCGGATACCGTTG	5095
QY	1138	GGCACCATCTTCGATGCTTCCCGAGATGAGAGTGAAGAACTCCCGCTTTTGGATAC	1197
DB	5094	GCGCGCTTCTTCGCGCGCTTCCCGGATATGCGGCTGACCGGCGCGGTAGAGCGCGG	5035
QY	1198	CACCCCGCTTCCG 1211	
DB	5034	CGGCGCGCTTCCG 5021	

Sat Oct 4 14:14:10 2003

us-10-014-717-1_copy_62369_63628.rge

Page 18

Search completed: October 4, 2003, 03:31:07
Job time : 4517.41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:01 ; Search time 8424.03 Seconds
(without alignments)
12307.996 Million cell updates/sec

Title: US-10-014-717-1_COPY_7610_11875

Perfect score: 4266
Sequence: 1 gtcggcgatcgcccatcga.....gggaatcatcgcctatga 4266

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	211	4.9	871 29	B2566693 pac82-164
2	188	4.4	753 29	B2561977 pac82-164
3	180.2	4.2	979 29	B2568491 pac82-164
4	165	3.9	1303 29	B2577421 msh2_5402

5	129.4	3.0	865 29	B2566683 pac82-164
6	128.2	3.0	679 29	B2562208 pac82-164
7	126	3.0	1348 29	B2569008 pac82-164
8	123.4	2.9	988 13	BX384927 BX384927
9	123	2.9	4176 11	BC021544 Homo sapi
10	123	2.9	4185 11	BC014631 Homo sapi
11	118.8	2.8	1061 13	BQ069153 AGENCOURT
12	118.2	2.8	1201 13	BX32425 BX32425
13	114.6	2.7	1798 29	AG171124 Pan trogl
14	114.4	2.7	599 29	B2560333 pac82-164
15	113.4	2.7	899 13	BUS28700 AGENCOURT
16	112.8	2.6	1798 29	AG171124 Pan trogl
17	110	2.6	876 29	B2573950 msh2_3439
18	108	2.5	1101 29	AG039543 Pan trogl
19	107.8	2.5	1196 13	BQ069593 AGENCOURT
20	107.8	2.5	1198 29	CC190008 CH261-138
21	107	2.5	1438 29	B2557931 pac81-60
22	106.6	2.5	1785 29	CC219595 CH261-62D
23	105.6	2.5	1281 12	BG852363 1024034A0
24	105.6	2.5	1695 29	CC290874 CH261-172
25	105	2.5	499 12	BM62671 17006874
26	105	2.5	577 12	BM646821 17006873
27	104.6	2.5	1299 29	AG039481 Pan trogl
28	104.4	2.4	1610 29	B2569386 pac82-164
29	104.2	2.4	690 10	BB865539 BB865539
30	104	2.4	692 12	BM652119 17006873
31	103.6	2.4	1189 29	AG032118 Pan trogl
32	103.4	2.4	968 13	BUS57094 B2562999
33	102.8	2.4	1090 29	B2562999 pac82-164
34	102.6	2.4	625 12	BM091001 B2562999
35	102.2	2.4	766 10	BE293934 B2562999
36	102.2	2.4	1651 12	BG809816 msh2-164
37	102	2.4	600 29	B2560280 pac82-164
38	102	2.4	928 13	BQ957144 AGENCOURT
39	101.6	2.4	699 10	BF250997 B2560280
40	101.4	2.4	1542 29	AG032943 Pan trogl
41	101	2.4	598 29	B2560289 pac82-164
42	100	2.3	704 28	AQ449661 50003A11
43	99.8	2.3	731 10	BG698437 602658472
44	99.8	2.3	1242 12	BM911414 AGENCOURT
45	99.8	2.3	1350 12	BM450237 AGENCOURT

ALIGNMENTS

RESULT 1
B2566693/c
LOCUS pac82-164 6525.xl pac82-164 Pseudomonas aeruginosa genomic clone
DEFINITION pac82-164_6525, genomic survey sequence.
ACCESSION B2566693
VERSION B2566693.1 GI:27196732
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 871)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: Shotgun.
LcLocation/Qualifiers

FEATURES

```

source
1. 871
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164.6525"
/clone_lib="pac2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."

BASE COUNT      165 a      294 c      269 g      140 t      3 others
ORIGIN

Query Match      4.9%; Score 211; DB 29; Length 871;
Best Local Similarity 57.6%; Pred. No. 2.5e-27;
Matches 396; Conservative 0; Mismatches 290; Indels 1; Gaps 1;

QY 221 CTTTCCTGAGCAGCAGCTGCTGCTGCAAGCCTCTCTTCTTCGAGCATCTCGCTCGAAG 280
DB 768 CTTTCCTGATGAGGTGGAAGGCTTCGAGC-GCGTTCTTCGGGATATCCGCTTCGAGG 710
QY 281 CGCTCGGATGAGACCCCTGCAATCGACTCTTCGTGAGGTGTGTGGAGGCGCTGAGAG 340
DB 709 CTGCGGAATGAAACCCCGCAGCGGCTGCTGAGAGTCTGTGGGAATGATGAGAAC 650
QY 341 AGCGCGCATGCTCCATCGGCGCTGTGTGTGTAACGGAACGGAGTTCATCGGGATG 400
DB 649 GGGCGGGGCTGTGCTGTGAACGCTGAAGGTTCGCCGACTGCGGTCTACGTGGCATGC 590
QY 401 GCCCGTCCGATATGAGGCGCGCTGCGCAAGCGACGCGCTCGGAGATCGACGCTC 460
DB 589 TGGGATGATTAACCTTACCTGCAATGCGCGGAGGCGGATATCCCGGATCAATCAT 530
QY 461 ATGGGCGCTGGGAGCATGCCAGCGTGTGAGCGGCGGAGATCTGTATGCTTCGCGGC 520
DB 529 TCTATGCGGCTGCGCAAGAGTTCAGCTTGTGCGCGCGCGGCGCATGCGCTATCACTTGGCG 470
QY 521 TCGGAGGCGCGTGTGTGCGGTGAGATACGAGCCATTCGTCTGCTGTGTGCGGTTCATC 580
DB 469 TCGAGGCTCGGCGATGCGGTGAGCAGCGCTGCTCTCTTGTGCTGTGTGCGATTCACC 410
QY 581 TGGCGCTGTGAGAGCTTTCGCTCGCGGAGATGCTTCAAGCGCCCTGTGTGTGAGTATCG 640
DB 409 TGGGTTGTGCGGCTTTCGAGAAATCGGAGGCGGAGCCTGCGATAGCGCGTGGGTCAATC 350
QY 641 TGAATGTTGTGCGGAGACCTTCGTGTGCTCTCGAAGACCGGCGCTGTGCGGAGAGC 700
DB 349 TGAATGTTGTGCGGAGACTATCTCATGAGCCAGATCAAGGCGGATTTGCGCGAGCG 290
QY 701 GTGCGTGCAGAGCATTTTGGGCGGAGCGGATGGTTCGAGAGGCGGAAGGTTGCGCG 760
DB 289 GAGCGCTGCGGAGCTTCGAGCGCGCGCGGATGCAATGTCGCGGCGGAAGGAGGTGCG 230
QY 761 TCGGTGCTCTCAAGCGGCTCAGTGAGCGCGCGCGGAGCGGATCGGATATTTGGCGTGA 820
DB 229 CGTTGTGCTCAAGCGCTCTGAGAGAGCTCTGAGCGGAGCGGAGCCGATCTGTGGGTGA 170
QY 821 TTTCGAGATTCGCGATCAATCAAGAGGTGCGAGAGCGGCTGTGACCTGTGCGGAAGGGA 880
DB 169 TTTCGTGCTGCGGATCAACAGAGAGCGGAGCGCGCGGAGAGCGGATCTGTGCGGAGGCA 110
QY 881 GCTCCCAAGAAATCGTGTGAAGAGG 907
DB 109 ATGCCAGGCGGCTGTGGGCTGCGAG 83

```

```

RESULT 2
BZ561977      753 bp      DNA      linear      GSS 17-DEC-2002
LOCUS         pac2-164.3530.x1 pac2-164 Pseudomonas aeruginosa genomic clone
DEFINITION   BZ561977
ACCESSION   BZ561977
VERSION     BZ561977.1 GI:27183081
KEYWORDS    GSS.

```

```

SOURCE
ORGANISM      Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE     1 (bases 1 to 753)
AUTHORS       Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
              Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE         Whole-Genome-Sequence variation among multiple isolates of
              Pseudomonas aeruginosa library
JOURNAL       J. Bacteriol., (2002) In press
COMMENT       Contact: Chris K. Raymond
              Genome Center
              University of Washington
              Box 352145, Seattle, WA 98105-2145, USA
              Tel: 2062216954
              Fax: 2066857244
              Email: ckraymond@u.washington.edu
              Class: Shotgun.
FEATURES
source
1. 753
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164.3530"
/clone_lib="pac2-164"
/notes="clinical isolate 2-164 whole genomic shotgun
library."

BASE COUNT      127 a      250 c      243 g      133 t
ORIGIN

Query Match      4.4%; Score 188; DB 29; Length 753;
Best Local Similarity 60.3%; Pred. No. 2.7e-23;
Matches 311; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 633 GGTATGCTGATGTTGTGTGCGGAGCAGCCCTGTGTGCTCTCGAAGACCGGCGCTTGGC 692
DB 95 GCTTCCCTGATCCTTTGCGCCTATGTTGTGCTGTGCTGATGAGAAATTGCGGCGGTGGC 154
QY 693 CAGGAGCGGTGCTGCAAGGATTTTCGCGGAGCGGATGAGTTTGGACGAGCGGAAG 752
DB 155 CAGGATGAGGCGTTGCAAGACCTTTCAGCGCGCGCGGCGGCTATGCGCGGAGAGGG 214
QY 753 GTGCGGCGGTGTGCTTCAAGCGGCTGAGTGAAGCGCGCGGCGGAGCGGATTCGATATT 812
DB 215 CTGCTGTGTGCTGTTCTCAAGCGCTGTGCGGAGCGGAGCGTGAACGGAATCCGTTCT 274
QY 813 GCGGTGATTCAGAGATCCGATCAATCAGAGCGGTGCGAGCGGCTTGAACCGTCC 872
DB 275 CCGGATGATCAGGTTAGCGGCTTCAACAGAGCGGCTTCAGCGGAGATCAAGGTTCC 334
QY 873 GAAAGGAGCTTCCAGAAATGTGTGTAAGAGCGGCTTGGCGAGCGAGGCTGCGCGC 932
DB 335 GAAATTCACCGCCGAGAGCGGTATGCGCGCTGCTTGGCGCAATGCGGAGCCAGAGC 394
QY 933 GTTTCGTGTGTTATGTGAGGAGCAAGGAGCGGAGCGGAGCGGATGAGCCCATGA 992
DB 395 GAGCGAGTCACTACATGAGGAGCAATGTGACGCGGAGCTGCGGAGCGGAGTGA 454
QY 993 AATCCAGCTCTGATGCGGTATACGCGCTCGGCGGAGATGTGCGACGCGCTGCTGAT 1052
DB 455 ACTGCGCGGCTGATGCGGTGCTGCGGCGGAGCGCGGCTGACAGAGCGGCTGCTGCT 514
QY 1053 CGGTCGTGAAGACCAACTTGGCGATCTGAGTATGCTGTGGGATCACTGGGCTGCT 1112
DB 515 GGGATGATTAAGGCAATCGGCACTGAGCGCAATGTGCGCGGCGGCTGCGGAGCTGCG 574
QY 1113 GAAGTGTGCTTGTCCCTTCAGACGAGGAGATTCC 1148
DB 575 GAAGATCTGTGCTGCTTGAACAGACCAAGATTGCC 610

```

RESULT 3

BZ568491/c 979 bp DNA linear GSS 17-DEC-2002
LOCUS BZ568491
DEFINITION pacs2-164_7558.v2 pacs2-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ568491
VERSION BZ568491.1 GI:27201527
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 979)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
CONTACT: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source 1..979
location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_7558"
/note="Clinical isolate 2-164 whole genomic shotgun
library."
BASE COUNT 187 a 324 c 299 g 164 t 5 others
ORIGIN
Query Match 4.2% Score 180.2; DB 29; Length 979;
Best Local Similarity 55.1%; Pred. No. 6.5e-22;
Matches 372; Conservative 0; Mismatches 297; Indels 6; Gaps 1;
OY 186 CGATGCCCGGGGAGAGACGCCGCTTACGCGCATCTTCTGACGACGTAGCTGCTT 245
DB 753 CCAACCCGGTCAACACCCCGGGCCATGCAAGCTTGTGACGGGTGAAAGGGGTT 694
OY 246 CGAGCCCTCTTCTTGGGATCTCGCTTGGCAAGCGCTGGGATGACCTGCACATCG 305
DB 693 CACCAAGNACTTTTCCGATCTCCGACCGAAGCGCAAGCATGATCGGCAACG 634
OY 306 ACTCTGTGAGGTGTCTGGAGGCGCTGAGAAACCGCGATCGTCCATCGGCCCT 365
DB 633 CTGTCTGATGTGCTTGGGAGCGCTGAGGAGGAGGTATTCGCGCATGACTT 574
OY 366 GTCGCTACGGAACGGAAGTGTTCATCGGATCGGCGCTCGAATATGAGCGCGCT 425
DB 573 GCGGGAAGCGACGTGGTGTTCGTTGGGCGCAATGCCAGCTAGAGGACGTTT 514
OY 426 GCCCAAGCAGCGGCTCGCAGAGATGACGCTATGCGGGCTGGGAGCATGCCAG 485
DB 513 GCTGG-----CTGCGCGCGCGGATGACCGAATACGCTACCGGAGCGCTTCTC 460
OY 486 GTCGGAAGCGGAGCAATCTGTATGCTCGGGCTGGAGGAGCGGTGTGCGGGTGA 545
DB 459 GGGCATTTCCGGCCGACGTGCAATTTCTTGGCTTGGCGGTCTCCCTGACCGTGA 400
OY 546 TACGCGCTATTCTGCTGCTGTCGTCGCTTATCTGACCTGTGACGCTTGGCTGG 605
DB 399 CACGCGGTGCTGTATGCTGACGCGCTGTGCACTGCGCTGCAACAGCTGTAGCGG 340
OY 606 GGAATGTCACAGCGCTGCTGGGTGTGGGGTATCGCTGATTTGTGCGGAGCACCTGTGT 665
DB 339 GGAGTGCAGATTTGCTATGTCGCGGGGTGAACGTATCGCATCTCGGCAATCTTCA 280

OY 666 GTGCTCTCGAAGACCCGGGCGCTGSCCAGGACGATCTGCAAGGCAATTTTCGCGGA 725
DB 279 ATCATGGGCGCATCCGCTGCTATCGGCGGACGCACTGCAAGACCTTCGACGATCG 220
OY 726 GCGCATGAGGTTCCGACGAGGCGAAGGTCGCCGCTGCTGCTCTCAAGCGGCTCAGTCG 785
DB 219 TGCCGATGCTACCGCGCGCGGAGGCTGTGTGATCTTCAAGCGGCGGACG 160
OY 786 AGCCCGCGCGGACCGCATGCGATATTTGCGGCTGATTCGAGATCCGATCAATCAGCA 845
DB 159 GCGCGGCGGACCGCGACCGGCTCATCGCTTCTGCTGCGGCTCGGCTCAACATCA 100
OY 846 CGTGCGAGCAGCGCG 860
DB 99 CGGCGCTGCGCGG 85
RESULT 4
BZ577421 1303 bp DNA linear GSS 17-DEC-2002
LOCUS BZ577421
DEFINITION meh2_5402.x1 meh Pseudomonas aeruginosa genomic clone meh2_5402,
genomic survey sequence.
ACCESSION BZ577421
VERSION BZ577421.1 GI:27212482
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1303)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
CONTACT: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
source 1..1303
location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="meh2_5402"
/note="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT 278 a 362 c 359 g 295 t 9 others
ORIGIN
Query Match 3.9% Score 165; DB 29; Length 1303;
Best Local Similarity 57.9%; Pred. No. 3.1e-19;
Matches 307; Conservative 0; Mismatches 220; Indels 3; Gaps 1;
OY 285 GCGATGACCCCTCAATCGACCTCTTGTGAGGTGTGAGAGCGCTGAGAAAGC 344
DB 94 GGGATGAGCCCGCAACAGCAATCGCGCTGAAAGTTCCTGGAAAGCTGAGACATGC 153
OY 345 GCGATGCTTCATCGGCGCTGTCGCTGTAAGAAACGGAATGTTTCATCGGATCGGCC 404
DB 154 GGTATGACCGCTTTCCTCGATGCGCAAGCTGTGCTTATGAGGGGCGAGTAC 213
OY 405 GTCCGAATATGAGCGCGCTGCGCAAGCAAGCGCGCTGCGCAAGATTCACGCTCATAG 464
DB 214 CAAGACTATATGACGCTGCTGCCAGCAACTCGGAGGTGCTGACGTCAATGCGCA--- 270

OY		465	CGGGCTGGGAGACGATGCCACGCCTGGAGCGGGCCGAATTCTGTATGACCCTCGGGCTGCG	524
Db		271	CCAATTCTAACGGCAGACCAAGCTTTATGTCGGACGATATCGCTACACCTTACGCGGCA	330
OY		525	AGGCGCGTGCTGTGCGGGTGGATACGAGCCATTTCCTTCGCTGTGTGGCCGTTTCATCTGAC	584
Db		331	AGGCCCGGAGATGATGATTCGACACCTCTGCTGCTTCCTTGTGTGGTGGCTTCACCAAGC	390
OY		585	CTGTACAGACTTGGCCCTCCGGGAAATGCTCCACGGCCCTTGGTGTGGGGATTCGCTCAT	644
Db		391	CTGCCAAGGCTGCTGCGCACGGGGAAGCGACCTTACGCTGGCCGCGAGATGAACCTTAAT	450
OY		645	GTTTGC GCCAGACACCTTCGTGTGACTCTCCGAAGA CCGGGGCGCTGGCCAGAGGACGATCG	704
Db		451	CCTTTCCGCTCATGATGGTTTCCTGTCTGTATGACAAGATTGGCGGCGCTGCACAGCATGGCG	510
OY		705	CTGCAAGGCATTTTTCGGCGGAGACCCGATGGATTTCGACGAGGCGAAGGATGCGCCGCTGT	764
Db		511	TTCGAAGA CTTTCGACGCGCGCGNCGACGGCTTATGGCTCTGNAGAGGGGCTGGCTGTGCT	570
OY		765	GGTCTCTCAAGCGGCTTCATGTGAGCCCGCGCGGACGGCCGATCGGATATTGCG	814
Db		571	TGTTCTCAAGCGGCTTGGCGGACGCGGAGCGGATGATCGTGTGCTGGG	620
RESULT 5		BZ566683		
LOCUS		pacs2-164_6520_x1 pacs2-164_Pseudomonas_aeruginosa genomic clone	865 bp DNA linear GSS 17-DEC-2002	
DEFINITION		pacs2-164_6520, genomic survey sequence.		
ACCESSION		BZ566683		
VERSION		BZ566683.1 GI:27196700		
KEYWORDS		GSS.		
SOURCE		Pseudomonas aeruginosa		
ORGANISM		Pseudomonas aeruginosa		
REFERENCE		Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
AUTHORS		1 (bases 1 to 865) Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V. Whole-genome-sequence variation among multiple isolates of Pseudomonas aeruginosa library		
JOURNAL		J. Bacteriol. (2002) In press		
COMMENT		Contact: Chris X. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: cgraymond@u.washington.edu Class: Shotgun.		
FEATURES		Location/Qualifiers		
source		1..865		
		/organism="Pseudomonas aeruginosa"		
		/mol_type="genomic DNA"		
		/strain="2-164"		
		/db_xref="taxon:287"		
		/clone_lib="pacs2-164_6520"		
		/clone_1ib="pacs2-164"		
		/note="clinical isolate 2-164 whole genomic shotgun library."		
BASE COUNT		142 a 263 c 298 g 161 t	1 others	
ORIGIN				
Query Match		3.0% Score 129.4; DB 29; Length 865;		
Best Local Similarity		53.5%; Pred. No.5.2e-13;		
Matches		296; Conservative 0; Mismatches 251; Indels 6; Gaps 1		
OY		756	CGCCGTCGTGTCTCAAGCGGCTCAGTAGACCCGCGCGGACGGCGATCGGATATTGGC	815
Db		45	CGAGTTCACGGTATTCGATTAACGCTTGATTCGATTCCTGCAGCGCCGACCGCGTCATCGC	104
OY		816	GGTGATTGAGGATCCGCGATCAATCAGACGGTGCAGACGCGGTCTGAACCTTGCACGA	875

Db	105	CGTTCTGTGGGCTCGGCGGTCAACATGACGGGGCCCTGGCGGGGGCTCAAGATCCAA	164
Qy	876	CGGAGACTCCCAAGAAATCGTGTGAAACGGGGCCCTGGCGGACGAGGCTGCGCGCTC	935
Db	165	CGGACCGGCCCCAGAGAGGCGTTGATTTCGCGAAGCCCTGGCCAAACGCCAGCTGTGCGCGC	224
Qy	936	TTCCGTGGGCTTATGTTCGAGGACACACGGGACGGGACACACGCTTGATGCCCATCGAAAT	995
Db	225	GCGGTGGGCTATGTTCAGGCGCCCATTTGGGACCGGACCACTGTTCGGCGATCCGATCGAGTT	284
Qy	996	CCAAAGCTCTGATGACGGGTATACGGGCTTCGGCGGAGATGTCCGACACCGCTGTGATCGG	1055
Db	285	GCAAGGCCCT-----GGCCAAAGCGCTATTCAGAGTACTGAGGCGCGTTCCTTGGGCGGTGC	338
Qy	1056	GTCGGTGAAAGCAACACTTGGCCATCTCTAGTATGCGTGGGGATCACTGGGGCTGTGCA	1115
Db	339	ATCGGTGAAAGTAACTGATCGGACATCTGGAGCGCGCGCGGTATCGCTTGTTGATCA	398
Qy	1116	GATGCTGTTCCTTCCTTACGACAGGGGAGATTCTTGCGACCTCCACGCGAGGGCGGTGAA	1175
Db	399	GAGCTGCTGTGTGTGTCGAGCGTGGAGTATTCGGAGAGGCTGTTCACGCTCTTAA	458
Qy	1176	CCCCCGGATCTCATAGGGGTGATCTTTCGGCTGACCGCTGACGCGCGCCGAGACCGGTGGC	1235
Db	459	CAGCCGCTTCACTGTGACGAGCAAGAGTCTCCAGTGTCCGAGGCAATCAGCCCTGGGA	518
Qy	1236	GGACTGGAATACGCCGCGACGGCGGGGGGTAGCTCGTTCCGATGACGGGACCAACGC	1295
Db	519	GGTGCGACGCGGACAGACGATGCGGGGATCAGCGCTTCGGTTTCAACCGGTACCAATGT	578
Qy	1296	GCAGCTGTGTGCTG 1308	
Db	579	CCATGTCTGTGTG 591	
RESULT 6			
BZ562208/c	BZ562208	679 bp	DNA linear
LOCUS	pacs2-164_3644.x1	pacs2-164	Pseudomonas aeruginosa genomic clone
DEFINITION	pacs2-164_3644, genomic survey sequence.		
ACCESSION	BZ562208		
VERSION	BZ562208.1	GI:27183746	
KEYWORDS	GSS.		
SOURCE	Pseudomonas aeruginosa		
ORGANISM	Pseudomonas aeruginosa		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.		
AUTHORS	1 (bases 1 to 679)		
TITLE	Spencer, D. H., Raymond, C. K., Smith, E. E., Sims, E. E., Hastings, M., Burns, J. L., Kaul, R. and Olsen, M. V.		
JOURNAL	Whole-Genome-Sequence Variation among multiple Isolates of Pseudomonas aeruginosa library		
COMMENT	J. Bacteriol., (2002) In press		
	Contact: Chris K. Raymond		
	Genome Center		
	University of Washington		
	Box 352145, Seattle, WA 98105-2145, USA		
	Tel: 2062216954		
	Fax: 2066857244		
	Email: craymond@u.washington.edu		
	Class: Shotgun.		
FEATURES	Location/Qualifiers		
source	1..679		
	/organism="Pseudomonas aeruginosa"		
	/mol_type="genomic DNA"		
	/strain="2-164"		
	/db_xref="taxon:287"		
	/clone="pacs2-164_3644"		
	/clone_1ib="pacs2-164"		
	/note="clinal isolate 2-164 Whole genomic shotgun library"		
BASE COUNT	135 a 237 c 202 g 101 t 4 others		

ORIGIN

Query Match 3.0%; Score 128.2; DB 29; Length 679;
Best Local Similarity 59.3%; Pred. No. 8.3e-13;
Matches 239; Conservative 0; Mismatches 158; Indels 6; Gaps 1;

OY 1623 CGCCTTCTCTTACCGGACAGGGGCGCAGACGCTGGGACATGGCCCTGGCTGTACGA 1682
DB 515 CGCCTGCTGTCTCCCGCCAGGCGCTGCAGATCCGGGAATGGCCAGGACGCTGTACGA 456
OY 1683 TGTATGTCCTCGCTCCCGAGGCGCTTGACCTGTGCGTAGAGCTGTTCACAGAGACT 1742
DB 455 CAGCAGAGTCCGCTTCCTGTAGAGCAATGAGATGTCTGTGGCTTGGCGGAGATGG 396
OY 1743 CGACCGCGCGCTCCGAGAGTGTATGGGCGCAACCGGACAGCTGACGCGCGCTGCT 1802
DB 395 TGTGAGCTGTCTCGCGCTTGT-----CGAGTCGAGCGCGGAGGAGAGGCGCTTGT 342
OY 1803 CGACCAAGACGCTTACCCAGCGGCGCTGTTCACCTTGAATATGCGCTGCGCGCT 1862
DB 341 GCGCAGACAGCGCTATGCGAGCGCGGCGATTTTCGCGGAAATAGCGGCTGCGCGCT 282
OY 1863 GTGCGGCTGTGGGCTGTAGAGCCGAGTGTGCTCCGCGCATACATCGGTGAGCTGT 1922
DB 281 GTGATGCGCTGTGGGCGGTGTCCCGGCGCGCTGTTCGATCAAGTTTGTGGAATTTGG 222
OY 1923 GACTGCTGCGGTGGGCGGCGGTCTCTCGCTTGAAGAGCGGCTGTCTGTGGCTGGCG 1982
DB 221 GCGCGCTGTATCGCGGCGCGCTTACGCTGAGAGAGCGCGCTGAGAGTGTGAGGCGCG 162
OY 1983 CGGCGCGCTGATGACAGCGCTGCGCGCGCGCGCGCGATGCTG 2025
DB 161 CGGAGCGCTGGCGCCAGAACTGATGACCGCGGAGGGGTGATG 119

RESULT 7
B2569008/c 1348 bp DNA linear GSS 17-DEC-2002

LOCUS B2569008 1348 bp DNA linear GSS 17-DEC-2002
DEFINITION pac82-164.8195.x1 pac82-164 Pseudomonas aeruginosa genomic clone

ACCESSION B2569008
VERSION B2569008.1 GI:27202907

KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa

REFERENCE 1 (bases 1 to 1348)
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library

JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: ckraymond@u.washington.edu
Class: shotgun.

FEATURES
source 1. 1348
Location/Qualifiers

BASE COUNT 319 a 389 c 348 g 274 t 18 others
ORIGIN
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac82-164.8195"
/note="Clinical isolate 2-164 Whole genomic shotgun
library."

Query Match 3.0%; Score 126; DB 29; Length 1348;
Best Local Similarity 51.6%; Pred. No. 2.1e-12;
Matches 299; Conservative 0; Mismatches 279; Indels 1; Gaps 1;

OY 129 CGTGGGCGAGTCCCCCGGACCGCTGGAGTACAGACGCTGTGATCCCGACCCGA 188
DB 674 CGTCNCGGCTGCTCCCGCGGCGCTGGAATGGCG-AGCGATTTCACGACCTGTGTGCA 616
OY 189 TGCCCCGGGGAAGCGCCCGTTAGCGCGGCAATCTTCTCGAGCGACGATAGCTGTTGA 248
DB 615 GCGTCCGAGTACCACTATGCGNNNGAGAGGCGCTTCTTATAGAGTGAACGCTTGA 556
OY 249 CGCCTCTTCTTGGGCAATCTGCGCGGAGCGCTGCGAGTGAACCTTGACATGACT 308
DB 555 CGCGGCTTCTTGGGAAATCGCTTGAAGCTGCGAAATGACCTTCGACGAGCGCT 496
OY 309 CTGCTGAGGTGTGCTGGAGGCGCTGAGAGAGCGCGATCGCTCATGCGCGCTGT 368
DB 495 GCTCTGAGGTCGTCTGGAATGATGAAACGGGCGGCGCTGCTTGAACGCTGAA 436
OY 369 CGGTACGGAAGGAGGTTTCAATCGGATGCGCCCGCGAATATAGGCGCGCTGCC 428
DB 435 AGGTTCGCGCACTGCGGTCTAGTGGCATGCTGGGCAATGATTAATTACCTTGATGC 376
OY 429 GCAAGCGACGCGCTCCGAGAGTCAAGCTCATGCGGCGTGGGAGATGCCAGCGT 488
DB 375 GCGGAGGCGCGTATCCCGGATCAATTCATTCGCGCTGGCAAGATTCAGCTT 316
OY 489 CGGAGCGGCGGAATCTGTATGCGCTGCGGCTGCGAGGCGCTGTGTGCTGCGGTGATAC 548
DB 315 CGCGCGCGGCGCATGCGCTATCACTTGGCGGTGACACGCTCGGCGAGTGGGTGAGCAC 256
OY 549 GGCCTATTCGTCCTGCGGTGCGGCTCATCTGCGCTGTGAGAGCTTGCGCTCCGGGA 608
DB 255 GCGCTGCTCTTCTGCTGCGGTGCGGATTCACCTGAGTGTGTGCGCTTGCAGAAATGCGCA 196
OY 609 ATGCTCCAGCGCGCTGCTGCTGAGGTATCGCTATGTTGTGCGCGACACCTTGCTGTG 668
DB 195 GCGGAGCTGCGCATAGCGCTGCGCTGATCAATGATGCTGCGCCGAGCTGACTATTA 136
OY 669 GCTCTGGAAGACCGGCGCTGCGGAGGAGCGTGTGCTG 707
DB 135 CATGAGCGAGATCAAGGCGATTTGCGCGAGCGGACGCTG 97

RESULT 8
BX384927 988 bp mRNA linear EST 08-MAY-2003

LOCUS BX384927 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL004YA22 5-PRIME, mRNA sequence.

ACCESSION BX384927
VERSION BX384927.1 GI:30459273
KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 988)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 8574.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DL004B11Q1pkcluster=8574.r. Contact :
Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Oy	3476	ACGTGATCTCAACCTCGCTTCGCGGCGAGCTGATGACAAAGATTCAATCCTCGCAT	3535
Db	1944	ACCTGGTTTGAATCTCTTGGCGGAAGAACTGCAAGCCAGCCTGAGTGCTTGCTTA	2003
Oy	3536	CGCAACGCCCGGTTTGTGAGCTCGGCAAGCGCACTGTTACGCGGATTAACAGCTCGGG	3593
Db	2004	CGCAAGGTGCGCTTCCTGGAATAATGGCAAAATTGACCTTTCTCAGAACCAACCGCTCGGCA	2063
Oy	3596	TGCGGCGCTTCTGCGCAATCTCTTCCTCGCTGTGATGATCTCCGGGGATGATGCTCG	3655
Db	2064	TGGCTATCTTCTTGAAAGACGTGACATTCACACGGGCTCTGATGATGCGTTCTTCAACG	2123
Oy	3656	AGCGGCGGCGCGGGTCCGTGCGCTCTTGAAGAACTCTTGCGCTGATTCGCGGCGAGCG	3715
Db	2124	AGAGCAGTGTGACTGCGCGGGAAGGTGTGGGCGCTTGTCAAGGCCGGAATCCGGGATGGGG	2183
Oy	3716	TGTTTCAACCCCTCCGCCCATTCGAGAGCTCCGATCGCGCGGTGCGCGATTCGTTCCGGA	3775
Db	2184	TGTTACGGGCCCCCTCAAGTGCACGAGTGTTCATGAGGGCCAGGTGAGGAGCGCTTCCGCT	2243
Oy	3776	GCATGCGCGAGCGCGAG	3792
Db	2244	ACATGCCCCAAGGGAAG	2260

RESULT 10	BC014631	LOCUS	BC014631	4185 bp	mRNA	linear	HTC 12-SEP-2002
DEFINITION	BC014631	Homo sapiens, fatty acid synthase, clone IMAGE:3351370,	BC014631				mRNA.
ACCESSION	BC014631	GI:17939526					
VERSION	HTC.						
KEYWORDS							
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
AUTHORS	1 (bases 1 to 4185)						
TITLE	Strausberg,R.						
JOURNAL	Direct Submission						
	Submitted (24-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA						
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov						
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Rubin Laboratory CDNA library Arrayed by: the I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Institute for Systems Biology http://www.systemsbio.org contact: amadan@systemsbiology.org Anup Madan, Jessica Fahay, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting						
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.llnl.gov Series: IMAGE:3351370 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis						
SOURCE	This clone has the following problem: frame shifted. Location/Qualifiers 1..4185 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:3351370" /tissue_type="Eye, retinoblastoma" /clone_id="NIH_MGC_16" /lab_host="DH10B-R" /note="Vector: pOTB7"						

	BASE COUNT	704 a	1421 c	1324 g	736 t	
Query Match	2.9%;	Score 123;	DB 11;	Length 4185;		
Best Local Similarity	50.2%;	Pred. No. 7.8e-12;				
Matches 340;	Conservative 0;	Mismatches 325;	Indels 12;	Gaps 1;		
OY	3128	TTGGCTACCCACAGTCAACGCGGTGGCTGGCGTGCTGGCCCTCGGCCTTCAGGCGCTTCGCG	3187			
Db	1593	TGGCACACTCTGTCTCTGTCTGTCAACGGACATTCTCTGGGAATGTCCCTTCCAACTGGAAGC	1652			
OY	3188	CGATCGAAGCGGCCCGCATGCCCCTCGCTGACTGACGCGCATGTGACGCGCTCGACAAGA	3247			
Db	1653	TGGAGAGAGCGCGCTCGGTGTCTGTCTGTACAGACAGGCTTACTAGCGCTGTGGTGC	1712			
OY	3248	TAGCCCGCTTCAAGCCGGGGAGCGGGTGTCTGATTCATGCGGCGAACCGGCGGGTCTGCTC	3307			
Db	1713	GTVGGCGGGTGCGCCCGGGGAGACGTGTCTCATCTCGGGCTCGGCGCGCTGGGCC	1772			
OY	3308	TCGCCCGCGTTCAGATGGGGCGGACGACGTTGGAGCGAGTCCATGCGACGCGCGGACGC	3367			
Db	1773	AGGCGCCATGTGCCATGCCCCCTCAAGTGTGGCTGCCGCTTTCACACACGTGGGGTGG	1832			
OY	3368	CCGAGAAACGCGCTTACCTGGAGTTCGTGGCGCTGGCGATGTAG-----CG	3415			
Db	1833	CTGAGAGACGGGCGTACTCCAGGCGAGGTTCCCCAGCTGACAGCACAGCTTTCGCA	1892			
OY	3416	ATTCCCGCTGGACCGGTTGTGTGCCCACTGGCGCGCGTGGACGGGCGGCGAGGAGTAG	3475			
Db	1893	ACTCCCGGACACATCTTTCAGAGAGATGTGCTGTGGCACACGCGCGGAGGAGCGTTG	1952			
OY	3476	ACGTGTGCTCAACTCGCTTCGGGCGAGCTGATCGACAAGAGTTTCATCTCTCGCAT	3535			
Db	1953	ACCTGTCTTAATCTCTTGGCGGAGAGAGACGTGAGGCCAGGTGATGTCTTGCTA	2012			
OY	3536	CGACACGCGGTTGTGGAGCTCGGCAAGCGGCACTGTATCGCGAATPACAGCTCGGCC	3595			
Db	2013	CGACACGCTGCTTCTGTGAATAATGGCAAATTGCACTTTCACAAACCACCGCTCGGCA	2072			
OY	2596	TGGCGCGGTCCTGCGCAATCTCTCTTCTCGCTGTGGATCTCGCGGGAGTGAATGCTCG	3655			
Db	2073	TGGCTATCTTCTTAAGAACGTGACATTCACGCGGGGTCTACTAGATGCGTTCTTCAACG	2132			
OY	3656	AGCGGCGGCGCGGGGTCCGTGCGCTTGTGAGAGAGCTCTTCGGCGTGTTCGCGGACGCG	3715			
Db	2133	AGACACGTGCTGAATGGCGGGAGGTGTGGGCGCTTGTGCAAGCGCGGACATCGGGATGGGG	2192			
OY	3716	TGTTACCCCTTCCCGCCCATGCGACGCTCCCGATCGCCGTTGCGCGATGCTTTCGGGA	3775			
Db	2193	TGTATCGGCGCCCTCAAGTGCACGCGTGTTCATGTGGGCGCACAGTGTGAGGACCGCTTCCGCT	2252			
OY	3776	GCATGGGCGAGGCGCAG	3792			
Db	2253	ACATGGCCCAAGGAGAG	2269			

RESULT	11
BQ069153	
LOCUS	
DEFINITION	BQ069153 1061 bp mRNA linear EST-02-APR-2002 AGENCOCURT_6738542 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5805443 5' , mRNA sequence.
ACCESSION	BQ069153
VERSION	BQ069153.1 GI:19898199
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Chordata; Craniota; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. Mammary gland tissue from fetal brain.
REFERENCE	1 (bases 1 to 1061) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished
AUTHORS	Contact: Robert Strausberg, Ph.D.
TITLE	
JOURNAL	
COMMENT	

Email: c9apbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LLCM2039 row: h column: 04
 High quality sequence scop: 563.
 Location/Qualifiers

FEATURES

1. 1061
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5803443"
 /tissue_type="neuroblastoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1ib="NIH MGC 47"
 /note="Organ: brain; Vector: pORF7; Site 1: XhoI; Site 2:
 EcoRI; CDNA made by oligo-dt priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCAAGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH MGC Library."

BASE COUNT

219 a 333 c 316 g 193 t

ORIGIN

Query Match 2.8%; Score 118.8; DB 13; Length 1061;
 Best Local Similarity 51.1%; Pred. No. 3.8e-11;
 Matches 339; Conservative 0; Mismatches 312; Indels 13; Gaps 2;

3193 GAGGCGGCGCCATGCGCGCTCCGCTACCTGAGCGGCAATGCTCGACAGAAATAGCC 3252
 40 GAGGCGGCGCTCGTGTCTGCTGCTACAGACGCGCTACTACCGCTGTGCTGCTGGG 99
 3253 GCGCTTACGCGGCGGAGCGGGTGTGATTCATGCGGCAACGCGGGGTGCTGCTGCC 3312
 100 CCGGTCGCGCCCGGGGAGAGCGCTGCTCATCTCGGGCTCGGGCGCGCGCCAGGCC 159
 3313 GGGGTGCAATGGGCGGAGAGCGGAGCGGAGTGCATGCGACCGGCGCGGCGCCGAG 3372
 160 GCGATGCGATGCGCTTCACTGCTGCGCTGCGCGCTTTCACACCGTGGGTGG 219
 3373 AAACGCGCTTACCTGAGTGTGCTGCGCTGCGGTATGTAG-----CGATTTC 3420
 220 AAGCGGCGTACTCTCCAGGCGAGGTTCCCGCAGCTGAGACACCGAGTCCGCAATCC 279
 3421 CGCTCGGACCGGTTCTGCGCGCAGCTGCGCGCGGTGAGCGGCGGAGGAGTGAAGCTC 3480
 280 CCGGACACATCTTTCAGAGCAGATGTGTGGCACACCGGCGGAGGCGCTTGAACCTG 339
 3481 GTGCTCACTGCTGCTGCGGCGAGCTGATGCAAGAGTTTCAATCTCTGCGATGCGAC 3540
 340 GTCTTGAATCTCTTGGCGGAGAGAGAGTGCAGGCGAGGTGCTTGGCTAGCAGC 399
 3541 GGCCTGTTTGTGAGCTCGGCAAGCGGAGTGTTCAGCGGATTAACAGTCCGCGCTGCG 3600
 400 GGTGCTTCTGGAATTTGGCAATTGCACTTTTCAGAACACACCGGCTCGGCAATGGCT 459
 3601 CCGTTCGTCGCAATCTCTCTCTTCTGCTGTGTGATGCTCCGGGGAGTATGCTCGAGCG 3660
 460 ATCTTCTCAAGAACCTGACATTTCCACGGGGTCTCTAGATGCGTTCTTCAACGAGAC 519
 3661 CCGGCGGCGGCTCGGCGCTCTTGTGAGAGCTCTCTGCGCTGATGCGGCGAGCGGTTC 3720
 520 AGTGTGACTGTGGGGAGGTGTGGGCGCTTGTGCAAGCGGCGGCGGAGTGGGGTGA 579
 3721 ACCCTTCCCGCATCGGAGCGCTCCGATGCGCGT-GTTCGCGATGCGTTCCGAGCAT 3779

Db 580 GCGCCCTCAATGACGCGTGTTCATGCGGCGCCAGGTGAGAGACGCTTCGCTACAT 639
 Qy 3780 GCGGAGCGCGACATCTTGGAAAGCTGTACTACGCTGCGTGAACCGGAGGTCCAGAT 3839
 Db 640 GGGCCGAGGAGACCATTTGGCAAGTGTCTCGAGGTGCTTGGAGAGACCGGAGG 699
 Qy 3840 CCGT 3843
 Db 700 CAGT 703

RESULT 12

EX332425

EX332425 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens

CDNA clone CS0DC012101 5-PRIME, mRNA sequence.

ACCESSION BX332425
 VERSION BX332425.1 GI:30335143

KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 1201)
 Li W.B., Gruber C., Jesssee J. and Polayes D.
 Full-length CDNA libraries and normalization

Unpublished
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library: seqref@genoscope.cns.fr, a division of
 Invitrogen. This sequence belongs to sequence cluster 8574.r For
 more information about this cluster, see

<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DC012AE01Q1&cluster=8574.r>. Contact :
 Feng liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/Invitrogen Corporation 1600>
 Paraday Avenue Genoscope sequence ID : CS0DC012AE01Q1.

FEATURES

source
 1. 1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DC012Y101"
 /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
 /clone_1ib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo (dT)
 primer..Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT

237 a 340 c 362 g 207 t

ORIGIN

Query Match 2.8%; Score 118.2; DB 13; Length 1201;
 Best Local Similarity 52.5%; Pred. No. 4.9e-11;
 Matches 287; Conservative 0; Mismatches 253; Indels 7; Gaps 1;

226 CTGAGCGAGTACCTGCTTGCAGCGCTCTTCTTGGCATCTCGCTCGGAGCGCTG 285
 337 CTGAGAGACCTGTAGGTTGATGCTCTCTTCTTGGAGTCCACCCCAAGCAGCAGC 396
 286 CGGATGAGACCTGACATCTGCTTGTGAGGTGTCTGGAGGCGCTCGGAGAACCC 345
 397 ACGATGAGACCTGACATCTGCTTGTGAGGTGTCTGGAGGCGCTCGGAGAACCC 456
 346 GCGATGCTTCATCGGCGCTGCTCGTACGGAACGAGAGTGTTCATCGGATCGGCGCG 405
 457 GGCATCAACCGATTCATCTCGGAGAACACACATGCGCT-----GGGTGGGGGT 509
 406 TCCGAATAGAGCGCGCTGCGGAGCGAGCGGTTCGAGAGATGAGGCTCATGAGC 465
 510 GACGCTCTGAGACCTCGGAGGCGCTGAGCGAGACCCCGAGACACATGCTGGGCTAAGC 569

Db	1379	GCCCGCGGAGNCGCCGNCCTGGCGCGCGCGCCNCGCGCNCGAGCCGCGCGAGNC	1438
Qy	2850	CGAGTCGATGAGCCAGCGCTGCTTGATCACCTCGTGTTCGGGTCACGAGCGCGCGC	2909
Db	1439	CGGCGCGCNCNCGGGNCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCNCNCN	1498
Qy	2910	CCCTGTGTGGGCGAGGTGCGAGATCGCTGAGCGCGCGGGGCTCAGCTTCAATGATGT	2969
Db	1499	CCGCGCCCGCNCNCNCNCGNCGCGCGCGCNCNCGCGCNCNCNCNCGCGC	1558
Qy	12970	CCACCTGCGGCTGGGCATGTGTCCCGAGCATCTCGCGGAAAGCCCAACCTCCGCTGCT	3029
Db	1559	CGGCGGCNCN	1615
Qy	3030	GCTCGGAGGCGAGTGTGCGCGCGCGCATGTGCGC-CGTGGGCGAGGAGCGTGAACGGGCTCG	3088
Db	1616	CGCGCGCNCNCNCG	1675
Qy	3089	TGTGGGCGCAACCGGTCACTGCGCTTTCGGCGGAGCGTTTGACCCAGCTCAACAGT	3148
Db	1676	CGCGCNCCTGGCGCGCGCGCGCGCGCGCNCNCNCGCGGNCNCNCNCNCNC	1735
Qy	3149	CGGCTGCGCTGTGTGCTCGGCTCAAGGGGCTCTCGGCGATGAGGCGGCGCGCATGC	3208
Db	1736	GCGCGCGCGCGCGCGCNCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1795
Qy	3209	CC 3210	
Db	1796	GC 1797	

RESULT 14	
B2560333	
LOCUS	B2560333
DEFINITION	B2560333 599 bp DNA linear GSS 17-DEC-2002
	pac62-164_2364.x1 pac62-164 Pseudomonas aeruginosa genomic clone
	pac62-164_2364, genomic survey sequence.

VERSION	KEYWORDS
BZ560333.1	GI:27178168
GSS.	
SOURCE	ORGANISM
Pseudomonas aeruginosa	Pseudomonas aeruginosa
Pseudomonas aeruginosa	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE	AUTHORS
1 (bases 1 to 599)	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.T., Kaul,R. and Olsen,M.V.
TITLE	JOURNAL
Whole-Genome-Sequence Variation among multiple isolates of Pseudomonas aeruginosa library J. Bacteriol., (2002) In press	
COMMENT	Contact: Chris K. Raymond

University of Washington
Box 353145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

```

FEATURES
  SOURCE
    Location/Qualifiers
      1..599
        organism="Pseudomonas aeruginosa"
        mol_type="genomic DNA"
        strain="2-164"
        db_xref="taxon:287"
        clone="pacs2-164_2364"
        clone_1fb="pacs2-164"
        note="clinical isolate 2-164 whole genomic shotgun
library."

```

BASE COUNT	89 a	184 c	199 g	126 t	1 others
Query Match					
Best Local Similarity	2.7%	Score 114.4;	DB 29;	Length 599;	
	52.2%;	Pred. NO. 2.2e-10;			

	Matches	253;	Conservative	0;	Mismatches	232;	Indels	0;	Gaps	0;
OY	34	GATCCGATTGGGATCGTCGGAGCAGATTGCCGCTGCGCCGCTGCGCGATTCGATCTGAGC	93							
Db	111	GAACCCCTCGCAATCATCATCGACGAGCTTGCCTGTTTCCCGCGAGGTGCTGCGCCGAG	170							
OY	94	GGGTTCTGGAACGCTCCCTCGAAGGGCTGCGCGGACACGCTCGAGGAGAGTCCCGCGCAACGC	153							
Db	171	GCATTCTGGAACCTCTTCCTTCGTCGTGGGACCGATTTCGTGGGCGCGTCCCGCGAGCCGC	230							
OY	154	TGGGATGCACAGCGGTGTTGATCCCGAACCCCGATGCCCGGAGGAGACGCGCGTTACG	213							
Db	231	TGGGATGGCAGAGCGGATTCAAGACGCTGTGGGACCGCTCCCGGTACCACTATTCGCGG	290							
OY	214	CGCGCATCTTTCCTGAACGACGTAGCCTGCTTCGACGCTCTTCTTTCGCACTTCGCGCT	273							
Db	291	GAGGGGCGCTTCCTTGATGAGGTGAGCCGCTTCGACGCGCGTTTCTTCGGGAAATCCGCT	350							
OY	274	CGCGAACCGCTGCGGATGGAACCTGCACATCGACTCTTCTGAGAGGTGCTGGAGAGCG	333							
Db	351	TCCGAGGCTCGCGAAATGGAACCCCGACGACGCGCTGCTGCTGAGAGTCTGTGGAAATGC	410							
OY	334	CTGAGAGAAGCGCGGATGCTGCTTCATCGGCGCTCGTGGTATCGGAAACGGGAGTTCATC	393							
Db	411	ATGGAACGGGCGGGGCTCTGCTCTGAAACGCTGAAAGSTTCCGCACTGGGCTTACGTC	470							
OY	394	GGGATCGGCGCCGTCGGAATATGAGGCCGCGCTGCGGACGACGCGCGTCCGACAGATC	453							
Db	471	GGCATGTGCGGATGATTAACCTTACCTTCGATCGCGGAGCGCGGTATTCGCGGCATC	530							
OY	454	GACGCTCATGGCGGGCTGGGAGCATATCCAGGCTCGAAGCGGAGCCGAATCTGTAAGCC	513							
Db	531	AATCATCTTAATGCGGCTGGAGNAGGATTGAGCTTCGCGCGCGGCGCATTCGCTATCAC	590							
OY	514	CTGGG 518								
Db	591	CTTGG 595								

RESULT 15
B0528700
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

B0528700 899 bp mRNA linear EST 13-SEP-2002
AGNCNCOURT 10183164 NIH_MGC.101 Homo sapiens cDNA clone
IMAGE:6538675 5', mRNA_sequence.
B0528700
B0528700.1 .GI:22839141
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 899)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Straubeberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC

```

FEATURES
    source
        CUNA Library Preparation: Kuhn Laboratory
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
        DNA sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNL at:
        http://image.lnl.gov
        Plate: L1CM2703 row: j column: 19
        High quality sequence stop: 600.
        Location/Qualifiers
            1..899
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6538675"
                /tissue_type="epidermoid carcinoma, cell line"

```



```
/lab_host="DH10B (phage-resistant)"  
/clone_lib="NIH_MGC_101"  
/note="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:  
XhoI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCACGAG(G). Library constructed by Ling Hong in the  
Laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH MGC Library."
```

```
BASE COUNT      145 a      280 c      307 g      166 t      1 others  
ORIGIN
```

```
Query Match      2.7%; Score 113.4; DB 13; Length 899;  
Best Local Similarity 49.4%; Pred. No. 3.3e-10;  
Matches 358; Conservative 0; Mismatches 352; Indels 14; Gaps 2;
```

```
QY 3128 TTGCTATCCACGTCACCACTGCGGTGCGTGTGCTCTCGGCTTCAAGCGCTCTTGG 3187  
Db 46 TGGCCACTCTGTCTCTGTCTCCTGTCCCGGACTTCTCTGGATGTGCTTCCAATGSAAGC 105  
QY 3188 CGATCGAGGCGGCGCGCCATGCGCCGTGCTACTGAGCGGCAAGGTAAGCGGCTCGACAGAA 3247  
Db 106 TGGAGAGGCGGCGCTTGGGTGCTGTCTGTACAGACGCGCTTACTACGCGCTGTGGTGC 165  
QY 3248 TAGCCCGCTTCAGCGCGGAGGAGGAGTGTATCATGCGCGGCGGAGTGTGTC 3307  
Db 166 GTGGGCGGGTGGCGCGGAGAGAGCTGTCTATCTCACTCGGGCTGGGCGGCGTGGGCC 225  
QY 3308 TCGCCGCGGTGCACTGGGCGGAGCACTGGGAGCCGAGGTCATGCGAGCGCGGACGC 3367  
Db 226 AGGCGCGCATCGCCATCGCTTCAGTCTGGGCTGCGGCTTCACTACCGTGGGCTGCG 285  
QY 3368 CCGAGAAACGCGCCATCGTAGTGTGCGGCGTGGGCTGTATGTAG-----CG 3415  
Db 286 CTGAGAGCGGCGCTTCTTCCAGGTCAGGTTTCCCCAGCTGACAGACCAAGCTTGGCCA 345  
QY 3416 ATTCGCGTCCGACCGGTTCTGTCCGCACTGTGCGCGCGTGGACGCGCGCGGAGTAG 3475  
Db 346 ACTCCCGGAGACATCTTTCAGAGAGCATGTGTGTGACACGCGGCGGAGAGGCGTTG 405  
QY 3476 ACGTGTCTCACTCGCTTCGCGGCGAGCTGATGACAGAGTTTCAATCTCTGCGAT 3535  
Db 406 ACTGTGTCTTGAATCTCTTGGCGGAGAGAGCTGCAAGCCAGGCTGAGTGTGGCTA 465  
QY 3536 CGCAGCGCGGTTGTGAGCTCGGCAAGCGGACTGTACGGGATACCAAGCTCGGCG 3595  
Db 466 CGCAGGCTCGCTTCTGGAATGGCAATTCGACTTTCAGAACCAACCGCTCGGCA 525  
QY 3596 TCGGCGGCTTCGTCGCAATCTCTCTCTGTGTGATCTCCGCGGAGTAGATGCTCG 3655  
Db 526 TGGCTATCTTCTGAAGAACGTGACATTCACAGGGGTCTACTGATGCGTTTCAAGC 585  
QY 3656 AGCGGCGGCGCGGAGTCCGTGCGCTTGTGAGAGACTCTTGGCTGATCGCGGAGCG 3715  
Db 586 AGACAGATGTGACTGTGGGAGGTGTGGCGCTTGTGCAAGCGGCAATCCGGAGTGGG 645  
QY 3716 TGTTCACCCCTTCCCCCATTCGCGAGCTGCC--GATGCCCGTGTGCGCGATGCGTTCCG 3773  
Db 646 TGTACGGGCGCTTCAAGTGCAGGCTTCCCATGGGCGCAGGTGGAGAGAGCGCTTCCG 705  
QY 3774 GAGCATGGCGGAGCGCGAGCATTTTGGAGACTGTACTACAGCTGGGTGACCGGAGGT 3833  
Db 706 CTACTGATCCAGGGAGAAACATTGGCAAGTCCCGCTGAGGGGCTTGTCCGAGAGAA 765  
QY 3834 CCAG 3837  
Db 766 CCCG 769
```

THIS PAGE BLANK (USPTO)

Dh 7730 GCGGACACCGTGTGGGGAAGTCCCGCCGACAGCGTGGATGACAGCGGTGTGATCCC 7789
Qy 181 GACCCGATGCCCCGGGGGAAAGCGCCCGTTACGCGCGCATCTTTCTTGAGCGAGCTAGCC 240
Dh 7790 GACCCGATGCCCCGGGGGAAAGCGCCCGTTACGCGCGCATCTTTCTTGAGCGAGCTAGCC 7849
Qy 241 TGCCTGACGCGCTCTTCTTGGGCATCTCGCGGAAAGCGCTGGATGAGCCCTGCA 300
Dh 7850 TGCCTGACGCGCTCTTCTTGGGCATCTCGCGGAAAGCGCTGGATGAGCCCTGCA 7909
Qy 301 CATCGACTTTCGCTGGAAGTGTGCTGGAAGCGCTGGAAGCGCGCATGCTCATCG 360
Dh 7910 CATCGACTTTCGCTGGAAGTGTGCTGGAAGCGCGCTGGAAGCGCGCATGCTCATCG 7969
Qy 361 GCGCTGCTGCTGTAAGGAAACGGGAGTGTTCATCGGAGTTCGCGCCGTCGAAATGAGGCC 420
Dh 7970 GCGCTGCTGCTGTAAGGAAACGGGAGTGTTCATCGGAGTTCGCGCCGTCGAAATGAGGCC 8029
Qy 421 GCGCTGCGGCAAGCGAGCGCGCTCGGAGAGTGAAGCTCATGCGGGGCTGGGGAAGATG 480
Dh 8030 GCGCTGCGGCAAGCGAGCGCGCTCGGAGAGTGAAGCTCATGCGGGGCTGGGGAAGATG 8089
Qy 481 CCGACGCTGGAAGCGGCGGAAATCTGTATGCTTCGCGCTGCGAGGCGCTGTGTGCGG 540
Dh 8090 CCGACGCTGGAAGCGGCGGAAATCTGTATGCTTCGCGCTGCGAGGCGCTGTGTGCGG 8149
Qy 541 GTGATACGGCTTATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Dh 8150 GTGATACGGCTTATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8209
Qy 601 TCCGGGGAATGCTCCACGCGCCCTGGCTGGTGGGATATGCTGATGTTGCGCGAGCAC 660
Dh 8210 TCCGGGGAATGCTCCACGCGCCCTGGCTGGTGGGATATGCTGATGTTGCGCGAGCAC 8269
Qy 661 CTGCTGTGCTCTCGAAGACCCGCGCGCTGCGCAAGAGCGGTGCTGCAAGGCAATTTTCG 720
Dh 8270 CTGCTGTGCTCTCGAAGACCCGCGCGCTGCGCAAGAGCGGTGCTGCAAGGCAATTTTCG 8329
Qy 721 GCGAGGCGGATGCGTTGCGAGCAGGCGGAAGGCTGCGCTGTGTGCTCTCAAGCGGCTC 780
Dh 8330 GCGAGGCGGATGCGTTGCGAGCAGGCGGAAGGCTGCGCTGTGTGCTCTCAAGCGGCTC 8389
Qy 781 AGTGAAGCCCGCGCGGAGCGGCGATGGAATTTGGGCGTGAATTCGAGAGTCCGCAATCA 840
Dh 8390 AGTGAAGCCCGCGCGGAGCGGCGATGGAATTTGGGCGTGAATTCGAGAGTCCGCAATCA 8449
Qy 841 CACGACGCTGCGAGAGCGGTCTGAACCGTGCAGAAAGGAGGCTCCCAAGAAATCGTGTG 900
Dh 8450 CACGACGCTGCGAGAGCGGTCTGAACCGTGCAGAAAGGAGGCTCCCAAGAAATCGTGTG 8509
Qy 901 AAACGGGCGCTGGCGGAGCGGAGCTGCGCGCTCTTGGTGGGTTATGTGAGGCAAC 960
Dh 8510 AAACGGGCGCTGGCGGAGCGGAGCTGCGCGCTCTTGGTGGGTTATGTGAGGCAAC 8569
Qy 961 GGCAGGCGGACGACGCTTGTGTGACCCCATCGAAATTCAGAGCTCTGAAATGCGGTATACGC 1020
Dh 8570 GGCAGGCGGACGACGCTTGTGTGACCCCATCGAAATTCAGAGCTCTGAAATGCGGTATACGC 8629
Qy 1021 CTGCGGCGGAGATGTGCGCACCGCGCTGTGATCGGGTGGTGAAGACCAACCTTGGCCAT 1080
Dh 8630 CTGCGGCGGAGATGTGCGCACCGCGCTGTGATCGGGTGGTGAAGACCAACCTTGGCCAT 8689
Qy 1081 CTTGATGATGCTGCGGAGTCACTGCGCTGCTGAAGGTGTCTTTGCTCTTACGACGCGG 1140
Dh 8690 CTTGATGATGCTGCGGAGTCACTGCGCTGCTGAAGGTGTCTTTGCTCTTACGACGCGG 8749
Qy 1141 CAGATTCGTCGACCTCCACGCGGAGCGGAGCGCTGAAGCCCGGATCTCATGGGATATCT 1200
Dh 8750 CAGATTCGTCGACCTCCACGCGGAGCGGAGCGCTGAAGCCCGGATCTCATGGGATATCT 8809
Qy 1201 CCGCTGACCGTTCACGCGCGCGCGGACACCGTGGCGGAGCTGGAATACGCGCGAGCGGCG 1260
Dh 8810 CCGCTGACCGTTCACGCGCGCGCGGAGACCGTGGCGGAGCTGGAATACGCGCGAGCGGCG 8869

Qy 1261 GGGGTGACCTGCTTCGGCATGAGCGGAGCAACCGCGCACGTTGCTGGAAGAGCGGCGG 1320
Dh 8870 GGGGTGACCTGCTTCGGCATGAGCGGAGCAACCGCGCACGTTGCTGGAAGAGCGGCGG 8929
Qy 1321 GCGGCGAGTGCACACGCGCGCGCGGAGGAGCGAGAGCTGTGCTGTGCGCA 1380
Dh 8930 GCGGCGAGTGCACACGCGCGCGCGGAGGAGCGAGAGCTGTGCTGTGCGCA 8989
Qy 1381 AGGACCGGCTACGCCCTGAGATGACAGGCGCGCGGCTGCGGACCATCTGAGACTTAC 1440
Dh 8990 AGGACCGGCTACGCCCTGAGATGACAGGCGCGCGGCTGCGGACCATCTGAGACTTAC 9049
Qy 1441 CTTTCCGAGTGTCTGAGGAGATGAGGCTGTCAAGTCTGCGGAGAGCGCGGAGCGCATGAG 1500
Dh 9050 CTTTCCGAGTGTCTGAGGAGATGAGGCTGTCAAGTCTGCGGAGAGCGCGGAGCGCATGAG 9109
Qy 1501 CACCGGCTCGCGGATGAGCGGAGCGTGCAGGAGGAGGCTGCGGAGCGCTGAGAGCTGCG 1560
Dh 9110 CACCGGCTCGCGGATGAGCGGAGCGTGCAGGAGGAGGCTGCGGAGCGCTGAGAGCTGCG 9169
Qy 1561 GCGCAGGAGACAGCGTGCCTCGGTGCGGATGCTGCGGATGCTTCTCACGCGGCAAG 1620
Dh 9170 GCGCAGGAGACAGCGTGCCTCGGTGCGGATGCTGCGGATGCTTCTCACGCGGCAAG 9229
Qy 1621 CTGCGCTTCTCTTACACCGGACAGGAGGCGGAGACGCTGGGCAATGAGGCGGTGTAC 1680
Dh 9230 CTGCGCTTCTCTTACACCGGACAGGAGGCGGAGACGCTGGGCAATGAGGCGGTGTAC 9289
Qy 1681 GATGATGATGCTCGGCTTCCGAGAGCGTTCGACCTGTGCTGAGAGCTTTCACAGAGAG 1740
Dh 9290 GATGATGATGCTCGGCTTCCGAGAGCGTTCGACCTGTGCTGAGAGCTTTCACAGAGAG 9349
Qy 1741 CTGACCGGCGCTTCCGAGAGTGTGAGGCTGAGCCGACCGGACGCTGACGCGCGCTG 1800
Dh 9350 CTGACCGGCGCTTCCGAGAGTGTGAGGCTGAGCCGACCGGACGCTGACGCGCGCTG 9409
Qy 1801 CTGACCGAGAGCTTTCACCGAGCGGCGCTGTTCACCTTGAATGAGGCTGCGCGCG 1860
Dh 9410 CTGACCGAGAGCTTTCACCGAGCGGCGCTGTTCACCTTGAATGAGGCTGCGCGCG 9469
Qy 1861 CTGTGCGGCTGTGAGGCTGTGAGAGCGGAGTGTGCTGCGCGGCAATGAGATGAGCTG 1920
Dh 9470 CTGTGCGGCTGTGAGGCTGTGAGAGCGGAGTGTGCTGCGCGGCAATGAGATGAGCTG 9529
Qy 1921 GTGCTGCTGCTGCGGCGGCGGTGTCTGCTTGAAGACGCGGTGTCTGCTGCTGCTGCG 1980
Dh 9530 GTGCTGCTGCTGCGGCGGCGGTGTCTGCTTGAAGACGCGGTGTCTGCTGCTGCTGCG 9589
Qy 1981 CCGCGGCGGCTGATGACAGGCGCTGCGCGGCGGAGGATGATGATGATGATGATGATGATG 2040
Dh 9590 CCGCGGCGGCTGATGACAGGCGCTGCGCGGCGGAGGATGATGATGATGATGATGATGATG 9649
Qy 2041 GAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2100
Dh 9650 GAGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9709
Qy 2101 AAACGCTCCGAGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2160
Dh 9710 AAACGCTCCGAGACAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 9769
Qy 2161 GCGATGAGCGGCGGCGGCGGCGGAGCAAGGCGGCTGCGATGCGGATGCGGATGCGGATG 2220
Dh 9770 GCGATGAGCGGCGGCGGCGGCGGAGCAAGGCGGCTGCGATGCGGATGCGGATGCGGATG 9829
Qy 2221 CCGCTCATGAGCCCGGATGCTGAGAGCGTTCGCGCGGTGTGAGCGATGAGTACCGG 2280
Dh 9830 CCGCTCATGAGCCCGGATGCTGAGAGCGTTCGCGCGGTGTGAGCGATGAGTACCGG 9889
Qy 2281 CCGCGCTGATGCTGCTGCTGACGAACTTGAAGCGGAAAGCTTGCACAGAGAGTGAAGC 2340
Dh 9890 CCGCGCTGATGCTGCTGCTGACGAACTTGAAGCGGAAAGCTTGCACAGAGAGTGAAGC 9949

QY 2341 TCCTCGGAGCTATTGAGTGCACGACGCGAGAGGTGTGCTTCGCGATGAGTAAG 2400
Db 9950 TCCTCGGAGCTATTGAGTGCACGACGCGAGAGGTGTGCTTCGCGATGAGTAAG 10009
QY 2401 GCGCTGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2460
Db 10010 GCGCTGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10069
QY 2461 GCGCTGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2520
Db 10070 GCGCTGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10129
QY 2521 GCGCTGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2580
Db 10130 GCGCTGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10189
QY 2581 CTGCTCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2640
Db 10190 CTGCTCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10249
QY 2641 CTTTGGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2700
Db 10250 CTTTGGCAACGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10309
QY 2701 GCGCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2760
Db 10310 GCGCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10369
QY 2761 GACCGGCGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2820
Db 10370 GACCGGCGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10429
QY 2821 GAGCGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2880
Db 10430 GAGCGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10489
QY 2881 CTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 2940
Db 10490 CTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10549
QY 2941 GACCGGCGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3000
Db 10550 GACCGGCGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10609
QY 3001 CTGCGCGGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3060
Db 10610 CTGCGCGGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10669
QY 3061 GCGCTGCGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3120
Db 10670 GCGCTGCGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10729
QY 3121 GAGAGCTTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3180
Db 10730 GAGAGCTTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10789
QY 3181 CTCTGCGGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3240
Db 10790 CTCTGCGGCAACGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10849
QY 3241 GACAGATAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3300
Db 10850 GACAGATAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10909
QY 3301 GTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3360
Db 10910 GTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 10969
QY 3361 GGCAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3420
Db 10970 GGCAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11029
QY 3421 CGCTCGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3480

Db 11030 CGCTCGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11089
QY 3481 GTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3540
Db 11090 GTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11149
QY 3541 GGCAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3600
Db 11150 GGCAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11209
QY 3601 CGCTTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3660
Db 11210 CGCTTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11269
QY 3661 CGCGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3720
Db 11270 CGCGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11329
QY 3721 ACCCTCCCGCATCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3780
Db 11330 ACCCTCCCGCATCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11389
QY 3781 GCGAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3840
Db 11390 GCGAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11449
QY 3841 CGTATTCGACCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3900
Db 11450 CGTATTCGACCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11509
QY 3901 CTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 3960
Db 11510 CTCTGCTCTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11569
QY 3961 GTCTGCGAGGTGTGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 4020
Db 11570 GTCTGCGAGGTGTGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11629
QY 4021 CTCTGCGAGGTGTGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 4080
Db 11630 CTCTGCGAGGTGTGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11689
QY 4081 CTGAGCTGTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 4140
Db 11690 CTGAGCTGTGAGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11749
QY 4141 CTGTTGATGCTGTGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 4200
Db 11750 CTGTTGATGCTGTGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11809
QY 4201 GCAAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 4260
Db 11810 GCAAGCGCGCGCGTGTGCGGCACTTTCGTCAAGGTTCGGTAATCGACGTGTCTC 11869
QY 4261 CTATGA 4266
Db 11870 CTATGA 11875

RESULT 2
US-09-953-348-19
; Sequence 19, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingying
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; TITLE OF INVENTION: MITOCHONDRION BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953.348

```

: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: PCT/US00/06394
: PRIOR FILING DATE: 2000-03-10
: PRIOR APPLICATION NUMBER: 09/266965
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 153
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 19
: LENGTH: 3978
: TYPE: DNA
: ORGANISM: Streptomyces lavendulae
: US-09-953-348-19

```

Query Match	Score	DB	Length
Best Local Similarly	19.1%;	814.2;	3978;
Matches 1549; Conservative	57.6%;	Pred. No. 7.7e-170;	
	0;	Mismatches 1108;	Indels 30; Gaps 4;

OY	18	CGAGCGCGACGCGGAAAGATCCGATTCGATCGTCGAGGAGGTTGCGCTCGCGCGGTGG	77
Db	90	CGAGGAGCGCGGCCCCGCGAGCCCGTCGATCTGGGCAATGAGCGTCCGCTTCCCCGGGGA	149
OY	78	CGTGATCGATCTTGAGCGGGGTTCTGACGCTCTCTGAGGAGCTTCGCGCGACACCGTCGGAGCG	137
Db	150	CGTGGCATCGCCGGAGCGACTGTGGCAGCTGTGTCCGAGGGAGCGGAGCGCGTCAACGA	209
OY	138	AGTCCCCCGCGGAAC---GTTGGGATGCGACACCGGTGTTTATCCCCAGCCCCGATGGCCCC	194
Db	210	GTTTCCCCCGCGACCGGGGGCTGGGACGTGCAGCGCCGTCTACGACCCCGAGCCGGGACATCCC	269
OY	195	GGGGAAGACGCGCCGTTACGCGCGGCATCTTCTTGACGACGTAAGCTTGCTTGACGCGCTC	254
Db	270	GGGCAAGACGTACGCGCGCGCACGCGCGCTTCTTCAAGAGACGCGCGCGATTCGACGCGCG	329
OY	255	CTTCTTGGCGATCTCGCTTCGCGAAGCGCTGCGGATGGAACCTGTGACATGTCTTGTGCT	314
Db	330	CTTCTTGGCGATCACCGCCGCGAGCGCTCGCCATGGAACCGCAGCAGCCGATGATCAT	389
OY	315	GGAGGTGTGTGGGAGGCGCTGAGAGACGCGCGATGCTCATTCGCGCGCTCGTCGTATAC	374
Db	390	GGAGGTCTCTGGGAGGCGCTTGAGAGGCGGGGCTTCAACGCGACACCTTCGCGGGCGA	449
OY	375	GGAAACGGGAGTGTTCATCGGATCGGCGCCGTCGGAATATGAGGCGCGCTCGCGCAAGC	434
Db	450	GGAAGTGTGGGTCTTGTGTGGCTTCCAAACGACGACGATACGATGATCAACGTGTCTGACGC	509
OY	435	GACGCGGTCCGAGAGATGACGCTCATGCGGGCTGGGACGATGCTCAGCGTGGAGC	494
Db	510	GCGGAGGTGTGCGCGAGGGCTTC-----ATCGGAGACCGGCAACTCCGCAACATCTCTC	563
OY	495	GGGCGGAATCGTATGACCCTCGGGCTGCGAGGGCGGTGTGTGCGGAGTGAATGAGGCTA	554
Db	564	CGGCGCGGTGCTTACACTTTCGACTTCGAGGGCCCGGCGGTGTCTGTGACATCGGCTTG	623
OY	555	TTGCTCTCGCTGATGAGCGTTTCATCTTGACCTGTCAAGCTTGGCTTCGCGGAGATGCTC	614
Db	624	CTCTCTCGCTGTGTGCGCTGCACTGGCGCGGAGTCCCTGCGGACGAGGGGAGTGTCTC	683
OY	615	CAGGCGCTTGGCTGTGGGGTATCGCTGATGTTGTGTGCGCGACGACCTCTGTGTGGCTCTC	674
Db	684	CTGGGCGCTTGGCGGGGCGCGACGATGTATGCGCACGCGGACCGCTTCACTGATGATTGACG	743
OY	675	GAAAGACCGGGAGCTGAGCGCAGGGGACGATGCGCTGCAAGGACATTTTCGCGGAGCGCGATGG	744
Db	744	CGCGCAGCGGGGCGTGTGCCCCCGACGCGCGCTGTGCAAGTCTTCTTCGCGGACCGCGCACGG	803
OY	735	GTTTCGAGCAGGCGGAAAGGTGCGCGCTGTGTCTCTCAAGCGGCTCAGTGGAGCCGCGC	794
Db	804	CACCACTGTGTCGAGGGGCGCGGCGGTGTCTGTCTGGACCGGCGCTTCGAGCGCCGCGCG	863
OY	795	GAGCGCGATCGGATATTTGGCGGTGATTTGAGAGATCGCGCATCAATACAGCGGTGGAG	854
Db	864	CTTGAGCTACCCCGTGCACGCGGTATCTCGGGGCGCGCGCTGTCAACAGAGCGCGCGAG	923

QY	855	CAGCGGCTCTGACCGGCGGAGACCTCCCAAGAAATCGTGTGAAACGGGCGCTGGC	914
Db	924	CGCGGCGCTTGACCGGCGCCCAACGACCGGCGGCACAGCGGGTGATCCGCGACGACTGGC	983
QY	915	GGACGACAGGCTGCGCGCGCTCTTCGTGGGTATTATGTGAGGCACACGGCACCGGACGAC	974
Db	984	CAACGCCACGGGTGACCGGCGGACAGCGGTGACGGCGGTCCAGGCGACACCGGACCGGACCCC	1043
QY	975	GCTTGTGTACCCCATCGAATCCAAAGCTCTGAATGCGGTATATACGGGCGCTCGGGCCAGATGT	1034
Db	1044	GCTGGGCGCACCCGATCGAGCGCCAGGCGCTCTCTCGCACCTTACGGGCGGGCCCGGCGGCA	1103
QY	1035	CGCCACGCGCGGTGTGATCGGGTCCGTGAACACCAACTTGGCCATCTTGATATGCGTC	1094
Db	1104	GGGACGCGCGGTGTGGCTGGGCTGGCTGAATCGAACCTTGGGCCACACCGATCCGCGGC	1163
QY	1095	GGGGATCATGTGGGCTCTGAAAGTGCTTTGTCCCTTTACGACCGGCGAGATCTTCCGCGCA	1154
Db	1164	CGCGCGGGCGCGGTGCATCAAGATGTGTATGGTCCATGTGGGCGACGGGAAGCTGGCCCGGAC	1223
QY	1155	CGTCACGCGGAGGCGGTGAACCCCGGCACTGCAATGGGGGTATCTTGGGCTGACCGGTAC	1214
Db	1224	GCTGACCTTACGAGGCCACCCCGCGGTGACTGTGTCGCGGTGATCATCGGCTGCT	1283
QY	1215	GCGCGCGCGGACACCTGTGGCGGACTGGAATACGCGCGGACGGCGGGGGTGAAGCTGCT	1274
Db	1284	GACCGGAGCGCCAGGACTGGCGCGGACACCGGACACCGCGCGGTGCGCGCTGTGCTCTT	1343
QY	1275	CGGCAATGAGGGGGAGCCAAACGCGGCAAGTGTACTGTGAAGAGGCGCGCGCGGACAGTGC	1334
Db	1344	CGGCTTACGGCGACCAACCGCCCATGTGTATCTTGAAGGGCCCGCGCGCGGAGAGGACAC	1403
QY	1335	ACCGCG-----CGCGCGGAACGACCGGACAGCTGTGTGTCTGTC	1376
Db	1404	GGAAGCGCGCGTGCAGGACGTCCTCGTGCAGCGCGGGGCGCGGTGCGGTGGTGTCTC	1463
QY	1377	GGCAAGGACCGCGTACAGCTCTGATGACACAGCGCGCGGTGGCGCACTTGTGAGAC	1436
Db	1464	CGGCGCGAGGAGCGCGCGTGTCCGAGCGCAAGCGGCGCTTGGCGCGCCACTGACCGC	1523
QY	1457	CTAACCTTCCGAGTGTCTGGGCGATGTGTGGCTTCACTGTGCGACGACCGCGCAT	1496
Db	1524	GCGCGCGGACCTTGGCACCGGCGGACGTGGCCACCGCGTGGCCACACGCGGGGGCTT	1593
QY	1497	GAAGCACCGGCTCGCGGTGCGCGGACGTCGAGGAGGGGCTTGGCGGACGCTCTGACGC	1556
Db	1584	CGACCACCGGGCGCGCGTGTGTGGCGCGGACCGTGAAGGAACTGTGCGCGGCTCTCGCGGC	1643
QY	1557	TGCGGCGGAGGAGACGACGTCGCGCGGAGCGGTGCGGAGTATCGCCGATTTCTCACGCGG	1616
Db	1644	CTTGGCCACCGGAACCGCGCGCGGCTGTGTCACTCGGCGGACCCGCGCTCGGCGG	1703
QY	1617	CAAGCTCGCTTTCTTTCACCGGACAGGGGCGGACAGCTGTGGGATGTGGCGT	1676
Db	1704	CAAGGCGCGTCTCTTTCACCGGACAGGGGACAGCAGCAGCGCCGCGCATGTGGCGCGCAACT	1763
QY	1677	GTAAGATGTATGTGTCCGCTTCCCGAGAGGCTTGAAGCTGTGCGGTGAGGCTGTCAACCA	1736
Db	1764	GGCGGCTTACGACCGCTGTTCGCGGACCGCTTGAAGAGGTCTTGGCCGACGCTGACAC	1823
QY	1737	GAAGCTCACCGGCGGCTCCGCGAGAGTATGTGGGCGGACCGGCGACGCTGACGCGCGC	1796
Db	1824	GCACTCGGACCGGCGGCTGGCGGAGGTGTCTTGTGCGCGGAGCGGACGCGCGAGGCGGC	1883
QY	1797	GCTGCTCGACCAAGCTTTCACCCGACCGGCGCTGTTCACCTTTCGAATATGTGGCTGCG	1856
Db	1884	CTGTGCTGACACGACCGGCTTACACCGCCGCGCGCTGTTCGCGGTGACAGTGTGGCGTCT	1943
QY	1857	CGCGGTGTGGGCGGCTGTGGGTGTATAGCGCGGAGTGGTCCCGGCGCATACATCTGGGTGA	1916
Db	1944	GCGGCTGTGGAAGACTGTGGGCTTGGCGCCGCGCATGTCCGCGGCGCACTCGGCTGGCGA	2003
QY	1917	GCTGTGGCTGCTGCGTGTGGGCGGCGTTCCTGCTTGAAGACGCGGCTTCTGTGGCGC	1976

Db	2004	ACTGACCGCCGCCCTACGCGCCGCCGGGGTGTGGTCCGTCCGCGCAGCCGCTGCGCCCTGTGTGCG	2063
QY	1977	TGCGCGCGGCGCGCTTGATGCAGGCGCTGCGCGCGCGCGGCGATGGTGTGCATGCAGGC	2036
Db	2064	CGCCCGCGGCGGCTGCAGCCGAGCAGCTGCCCGCGGCGGCGCCATGGTCCGCGTGCAGGC	2123
QY	2037	GCCGAGGCGCGATGGCTGCTGCGGATGGCGCGCAGCAGCTGCATGTGCATGCCGC	2096
Db	2124	GACGAGGAGCAGAGGTGCGCGCCCACTCGCGCAGCGCGCCCGCGTGGACATCGCGCG	2183
QY	2097	GGTCAAGCTTCGGGACCAAGTGTGATCGCGGCGCGCGGCGACCCGCTGCATGGCATCGC	2156
Db	2184	CGTCAACGAGACGGGAGCGGTGTGTGTCTCGGCGCAGAGGCGCGCTGCACGAGCTGGC	2243
QY	2157	GGCGGCGATGGCGCGCGCGGCGGCGCGGACCAAGGCGCTTCACGCTTCGATGGCTTCA	2216
Db	2244	GCGGAGTGGGCGCGCCGCGCGCGCGGAGACCAAGAGGCTTGCGGCTGACGACGCTTTC	2303
QY	2217	CTCACCGCTCAAGGCGCCCGATGCTGAGAGCGCTTCGCGGATGGCGGAGTCCGTGAGCTA	2276
Db	2304	CTCGGCCACCTGGAGCGCATGACGAGGCGTTCCGCGAGGTCCGACGAGGGGTGTCTTA	2363
QY	2277	CCGCGCGGCGGTGCATCGTCTGTGTGAGCAATCTGAGCGGAGAGGCTTGCACAGACGAGT	2336
Db	2364	CAGCGCGCGCTCCCTCCGCGGTGTCTTCCAGCGCTCACCGGCGCCCGCTCACCGACGAGCT	2423
QY	2337	GAGCTCGCGCGGCGCTATTGGGTGGCGCCAGCGCGAGAGGTGTGTGCGCTTCGCGATGAGT	2396
Db	2424	CCGCGAGGCGGGAGACATGGAGTGGCGGCAGTCCGGAGAGCGGTGGCGCTTCCAGACGCGGT	2483
QY	2397	GAAAGCGCTGCACGCGCGCGGTGGCGGCACTTGTGCAGGTCGAGTCCGAATGACGCT	2456
Db	2484	CCGCGCGCTTGGCGCAGCCGCGGAGCCACCGCTTCTGTGAGGTCCGAGCCCGCGCGGTCT	2543
QY	2457	GCTTGCGCGCTGTGTGCTGTGCTGCATGTCGCGAGCGCGCGCGC---GGCGTCTCGCATCGTC	2513
Db	2544	GAGGCGCGCGCAGCGCGAGTGCCTGTGCCGAGCGCGCCCGCGAGACGTTGCTCCCGTCT	2603
QY	2514	GCGCGCTTGGCGGTGACGAGCGCGCGACCTGTCTGAGGCGCTCGCGGCGCTTGTGGCGGT	2573
Db	2604	GCGCGCGCTGCGAGCGCCGAGACCGAGTCCGTGCTGACGCGCGCTCGCGCGAGGCCACAGAT	2663
QY	2574	CGGTGGCTGTGTCCTGTGGGCGCGGCTTTCCTTCAGAGGGGCGCGGAGTGGCGGTCTCC	2633
Db	2664	CGGCTCTTCGCGCGCGGTGGGAGCGCGCTGCTCCCAAGGCGCGGACGCGCGTGCATCTGCC	2723
QY	2634	CACGTACCGCTTGGCAGGCGGAGCGGTACTGTGATGCAGCAGAAAGCGC	2680
Db	2724	CACGTACCGCTTCCAGCGCGGCGCACTACTGTGCTGGCGGCAATGGCGC	2770

```

RESULT 3
US-10-267-255-19
: Sequence 19, Application US/10267255
: Publication NO. US20030124689A1
: GENERAL INFORMATION:
: APPLICANT: Sherman, D
: APPLICANT: Mao, Y
: APPLICANT: Varoglu, M
: APPLICANT: He, M
: APPLICANT: Sheldon, P
: TITLE OF INVENTION: Mlcomycin biosynthetic gene cluster
: FILE REFERENCE: 600,445C051
: CURRENT APPLICATION NUMBER: US/10/267,255
: CURRENT FILING DATE: 2002-10-09
: PRIOR APPLICATION NUMBER: US 09/266,965
: PRIOR FILING DATE: 1999-03-12
: PRIOR APPLICATION NUMBER: US 08/624,447
: PRIOR FILING DATE: 1996-08-19
: PRIOR APPLICATION NUMBER: PCT/US94/11279
: PRIOR FILING DATE: 1994-10-06
: PRIOR APPLICATION NUMBER: US 08/133,963

```

```

: PRIOR FILING DATE: 1993-10-07
: NUMBER OF SEQ ID NOS: 145
: SOFTWARE: FASTSEQ For Windows Version 3.0.
: SEQ ID NO 19
: LENGTH: 3978
: TYPE: DNA
: ORGANISM: Streptomyces lavendulae
: US-10-267-255-19

```

Query Match	19.1%;	Score 814.2;	DB 14;	Length 3978;
Best Local Similarity	57.6%;	Pred. No. 7.7e-170;		
Matches 1549;	Conservative 0;	Mismatches 1108;	Indels 30;	Gaps 4;

QY	18	CGAGCCGCGACGCCGAAATCCGATTGGGATTCGTGGAGCCGATTGCGCTTCGCTCCCGTGG	77
Db	90	CGAGAGCGCGGCCCGCGAGCCCGTCCGGATCTGTGGCATGGCGTCCGCTTCCCGGGGA	149
QY	78	CGTATCGATCTGAGCGGGATTCTTGGAGCTCTTGAGGGCTCGCGCGACACCTGTGGGC	137
Db	150	CGTGGCATCCGCGGACGACCTGTGGCAGCTGTGTCCGAGGGCCGGAGCGCGGTCAACGA	209
QY	138	AGTCCCGCGCGAAC---GCTGGATGAGAGCGCTGGTGTATCCCGACCCCGATGCCC	194
Db	210	GTTCCCGCGGACCGGGGGCTGGACGTTCGAGCCGTTCTAGAACCCCGAGCCGGCACCCC	269
QY	195	GGGAGAGCGCCCGTTACGCGCGCATCTTCTTGAGCGACGTAGCTTGCTTGAACGCTC	254
Db	270	GGGACGAGACGTACGCGCGCACGGCGGCTTCTCAAGAGACGCCCGGATTGACAGCGCG	329
QY	255	CTTCTTTGGGATCTCGGCTCGCGAAGGCGTGGGATGGACCTTCGACATCGACTTTCCT	314
Db	330	CTTCTTTGGGATCACGCCGCGCAGGGCTCGCATGATCCCGACGACGCCCATGATAT	389
QY	315	GGAAGTGTGCGGAGGCGCTGAGAAACGCGCGCATGCTTCATCGGCGCTCGTCCGTAC	374
Db	390	GGAGGTCTCTGGAGGGGCTTTCAGACAGGGGGGCTCGACGCGACACCTCGGGGGGA	449
QY	375	GGAACCGGAGTGTTCATCGGATCGGCGCGTCCGAATATGAGGCCCGGCGTCCGCAAGC	434
Db	450	GGAAGTGGGGGTCTTTCGTCCGCTTCCAAACAGACATCACTGTATCAACGTCTCGAGC	509
QY	435	GAGCGGCTCGGCAAGATCGACGCTCATGGCGGGCTGGGGACGATGCCAGCGTCGAGC	494
Db	510	GCGGAGCGTCCGCGAGGGCTTTC-----ATCGGACCGGCACTTCGCGCAGCATCTTTC	563
QY	495	GGGCGGATCTCGTATGCCCTCGGGCTGCAGAGGCGGTGTGTCCGCGTGAATACGGCTTA	554
Db	564	CGGCGCGGTGGCTACACCTTTCGGCTTGAAGGGCCGGCGGTGTCCGTGCACACGGCTG	623
QY	555	TTGCTCTCGTGTGGGCCGTTTCACTTGGCTGTGCAGAGCTTGGCGTCCGGGAAATGCTC	614
Db	624	CTCTCTCTCGTGTCCGCTGCACCTGGCGCGCGACATCTCTTGGCGGACGGGGAGTGTCTC	683
QY	615	CACGGCCCTGAGTGTGGGGTATGCTGATGTTTCCGCGAGCAACCTCGTGTGACTTTC	674
Db	684	CTTGGCCCTGGGGCGGGCGCGAGAGTATGGCCACGCCCAACCCCTTCAATCAGATTACG	743
QY	675	GAAAGCCCGGCGCTGGCCAGGAGCGGTCTGCTGCAAGCATTTTTCGCGGAGGCCGATGG	734
Db	744	CCGCGACACGGGGCTTGGCCCCCGACCGCGCTGCAATCTCTTCTCGCGAGCCGCGAGGG	803
QY	735	GTTTCGACAGAGCGAAGGTGTCCGCCGTCTGTGTCTTCAAGCGGGCTCAGTGAAGCCCGGC	794
Db	804	CACCACTGTCTGAGGGCGGGCGCGGTGTGTCTGAGCCCGGCTCTCGGAGCGCCCGCG	863
QY	795	GGAAGCGGATCGATATTGCGCGGTGATTCGAGATTCGCGCATCATCAACGAGTGCAG	854
Db	864	CTTGGGCTAACCCGTGACGGGGTCACTCGGGGGCAGGCGGTCAACACGAGCGGCGGAG	923
QY	855	CAGCGGTCTGACGTGCTCCGAACGGGAGCTCCAAAGAAATGTGTCTGAACCGGCGCTTGGC	914
Db	924	CGCGGGCTTACCGCGCCCAACGAGACGGGGCGCAACAGCGGGTATCTCGGCGAGGACTGGC	983

```
OY 915 GGAGCAGGCTGCGCGCGCTTTCGTTGAGTATGTCAGAGCAACGCGACGCGGACAC 974
Db 994 CAAGCAGGCTGATGACGCGACGACGCTCAACCGGTCAGAGCAACGCGACGCGACCCC 1043
OY 975 GCTTGTGACCCCATCGAAATCGAAGCTTGAATGCGGTATACCGGCTTCGCGAGATGT 1034
Db 1044 GCTGGGCGACCGGATCGAGGCGCCAGGCGCTCTCGCACCTACGCGGCGGCGCGGGA 1103
OY 1035 GCGCACCGCGCTGCTGATGCGGTGCGTGAAGACCAACTTGGCACTCTGATATGCTG 1094
Db 1104 GGGCAGGCGCGCTGTGGCTGGGCTCGCTGAAGTCGAACCTGGGCAACCCACCATCGCGCG 1163
OY 1095 GGGGATCACTGGGCTGTGAAGGTGTCTTTCCTTTCAGCAGGCGAGATTCCTGGCA 1154
Db 1184 CGGCGCGGCGCGGCTCATCAATGATGTATGATGCGGCACTGCGGAGACCTGCGCGCAC 1223
OY 1155 CTTCCAGCGCAGGCGCTGMAACCCCGGATCTCATAGGGGTGATCTTTCGCTGACGTCAC 1214
Db 1224 GCTGCACCTCAAGAGGCCACCCCGGCGGTGACCTGTGTCGCGCGGTACGTACGGCTGT 1283
OY 1215 GCGCGCGCGGACACCGGTGCGGACCTGCAATACCGCGGACGCGGCGGGGTGAGCTGTT 1274
Db 1284 GACCCAGGCGCAGAGCTGGCCGCGACACCGGACGCGCGCGCTGGCGCTCTGCTCTT 1343
OY 1275 CGGCAATGAGCGGAGCAACGCGCAGGTGTGCTGGAAGAGCGCGCGCGGCGAGCGTAC 1334
Db 1344 CGGCTCAAGCGGACCAACGCGCCCATGTGATCTTGAAGGCGCGCGCGCGCGAGAGCAC 1403
OY 1335 ACCGCG-----GCGCGGAGCGACCGGCGAGAGCTGCTGTCTGTC 1376
Db 1404 GAGCGCGCGCTGCGGAGCTCTCTCGGACGCGCGGCGCGCGCGCTGCGTGTGCTCTC 1463
OY 1377 GGCAGAGACCGGCTGACGCTTGAATGACAGCGCGCGCGCTGCGGACATCTGAGAC 1436
Db 1464 CGGCGCGAGCGGCGCGGCTGCGGACGCGCGCGCGCGCGCGCTGCGGCGCGCTCACTACCGC 1523
OY 1437 CTACCTTCGACGATGCTTGGGAGATGTGGCGTTCAGTCTGGGAGCAGAGCGCGAGCGCAT 1496
Db 1524 GGGCGCGACCTGCGACCGGCGGAGGTGGCACCGGCTGCGCACCAACCGGCGGCGCTT 1583
OY 1497 GAGAGACCGGCTGCGGCTGCGGCGGACGTCAGAGAGGGGCTGCGGCGAGCGCTGAGACG 1556
Db 1584 CGACACCGGCGCGCGGCTGCTGCGCGGCGGACGCTGAGGAACTGCTGCGCGGCGCTGCGCGC 1643
OY 1557 TGCGCGCGAGGAGACAGAGCTGCGCGCGGCGGCTGCGGAGTATCTCGGATCTCTACGCG 1616
Db 1644 CTTGGCCACCGGAAACCGCGCGCGCGCTGTGTCAACCGGCGGAGCCCGCGCTCGCGCGG 1703
OY 1617 CAAGCTCGCTTCTCTTCAACGAGCAGGCGGCGGACAGAGCTGGGATGGCGCGTGGCT 1676
Db 1704 CAGGCGCGCTTCTCTTCAACGAGCAGGCGGAGCAGCAGCGCGGCGATGGCGCGGACT 1763
OY 1677 GTACGATGTATGTTCGCGCTTCCGCGAGGCGTTCGACCTGTGCGAGGCTTTCACCA 1736
Db 1764 GCGGCGCTCAAGACAGCTGTGTTCGCGAGCGCTTGAAGCGGTGTGCGGCCACACTGACCG 1823
OY 1737 GAGAGCTCGACCGCGCGCTCCGCGAGGTATGTGGGCGGAAACGCGGCGAGCTGAGCGCGC 1796
Db 1824 GACCTTCGACCGCGCGCTGCGGAGGTGTGTTCGCGGAGCGGCAACCGCGGAGCGCGC 1883
OY 1797 GCTGCTCGACGACGAGCTTCAACGAGCGGCGGCTTCACTTCGATATGCGCTGCGC 1856
Db 1884 CTTGCTCGACGAGAGGCTTCAACGAGCGGCGGCTTTCGCGCTGAGGCTGCGCTGCT 1943
OY 1857 GCGCGCTGTGCGGTGCTGTGAGAGCGGAGTGTGCGCGGAGTGTGCGCGCATAGCATGCTGA 1916
Db 1944 GCGGCTGTGAGGAGCTGGGCGCTTGGCGCGCGCATGTGTGCGGCGCACTCGGTGCGGCA 2003
OY 1917 GCTGTGTGCTGCTGTGTGCGGCGGCTGTCTGCTTGAAGACGCGGCTTCTGTGTGCG 1976
Db 2004 ACTGACCGCGCTTACGCGCGCGGCGGTGTGTGCTGCGGACGCTGCGCGCTTGTGTGCG 2063
OY 1977 TGCGCGCGGCGCGCTGATGACAGCGCGCTGCGGCGGCGGCGGATGTGTGATGAGGCG 2036
```

```
Db 2064 CGCCGCGGCGCGGCTGACCCAGGCACTGCCCGCGGCGGCGCGCATGTGTGCTGCAAGC 2123
OY 2037 GCGGAGGCGCATGTGTGCTGTGCGGTGGCGCGCGCACGACGTCGTGTGATGTCGCGC 2096
Db 2124 GACCGAGGACGAGGTGCGCGCCAACTGCGCGAGCGCGCGCGCGCGGCGGATGCGCGC 2183
OY 2097 GGTAAAGCTTCGGAACGAGGTGTCATGCGGCGCGCGCGGCGGACCCGTCGATGTGATGCG 2156
Db 2184 GGTAAAGCTTCGGAACGAGGTGTCATGCGGCGCGCGCGGCGGAGCGCGCTGTCAGAGACTGGC 2243
OY 2157 GCGCGCATGTGCGCGCGCGCGCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2216
Db 2244 GCGGAGTGGGCGCGCGCGCGCGCGGAGACAGAGAGCTGTGGGCTGACGACGACGCTTCA 2303
OY 2217 CTACCGCTCATGCGCGCGGATGCTGAGGCGGTCGTGGCGGTGCGGCGGATGCTGAGCTA 2276
Db 2304 CTGCGCCACCTGAGCGCATGACCGAGCGGTCGCGGAGGTGCGACGAGGCGGTCTCTA 2363
OY 2277 CGGCGCGGCGGATGCTGCTGTCGATGACGATCTGAGCGGAGAGGCTTGCAGAGAGAGT 2336
Db 2364 CAGCGCGCTGCTCTCTCGGCTGTCTCAAGCTCAACGCGGCGCGCGCGCTGACGAGAGCT 2423
OY 2337 GAGCTGCGCGGATATGAGTGTGCGCACGCGCGGAGAGGTGTGCGCTTTCGCGGATGAGT 2396
Db 2424 CGGAGGCGCGGAGACACTGAGTGTGCGGACGTCGCGGAGAGCGTGTGCTTCAAGAGCGGT 2483
OY 2397 GAGGCGCTGACGCGGCGCGGTGCGGAGCACTGTGTGAGGTGCTGCTCCGAATGACGCT 2456
Db 2484 CGGCGCTTGTGCGGACCGCGGCGGCGCACCGGCTTCTGAGGAGTGTGCGGCGCGGCGTCT 2543
OY 2457 GCTGCGCTGTGTGCTGCTGCTGATGCGGAGCGCGCGCGC---GGGCGGCTGCGATGCTC 2513
Db 2544 GACGCGCGCGGACGCGGATGCTGCGGAGCGCGCGCGCGGAGAGTGTGCTGCTGCT 2603
OY 2514 GCGCGCTGGGAGTGAACGAGCGGCGGACGCTGCTGAGGCGGCTGCGGCGGCTTCTGAGCGCT 2573
Db 2604 GCGGCGCGGAGCGCGGACCGGAGTGTGCTGAGCGGCGGCTGCGGAGCGGCGGACGAT 2663
OY 2574 CGGTGCGCTGTCTTCTGCGCGGCTTCTTCTTCAAGGCGGCGGCGGCTGCGGCTGCC 2633
Db 2664 CGGCTTCTGCGCGGAGTGTGAGCGGCTGCTGCTGCTGCGGAGCGGCGGCGGCTGCGC 2723
OY 2664 CACGACCTTGTGAGCGGCGGAGCGGCTACTGATGAGACGAAAGCGG 2680
Db 2724 CACGACCTTGTGAGCGGCGGAGCGGCTACTGATGAGCGGCGGCTGCGGCGGCTGCGC 2770

RESULT 4
US-09-953-348-74
; Sequence 74, Application US/09953348
; Publication No. US20030134398A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, David. H
; APPLICANT: Mao, Yingling
; APPLICANT: Varoglu, Mustafa
; APPLICANT: He, Min
; APPLICANT: Sheldon, Paul
; TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
; FILE REFERENCE: 600.530US1
; CURRENT APPLICATION NUMBER: US/09/953,348
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: PCT/US00/06394
; CURRENT FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 12249
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-953-348-74
```

Query Match 19.1%; Score 814.2; DB 12; Length 12249;
Best Local Similarity 57.6%; Pred. No. 7e-170;
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

Qy 18 CGAGCGGCGACCGAAGATCCGATTGCGATGTCGAGCGAGATTGCCGTGCGCGGTG 77
Db 1077 CGAGGACCG 1136
Qy 78 CGTATGATGATGACCGCGGTTCTGGAAGCTCTCGAGGGGCTCGCGCGACACCGTGGGCG 137
Db 1137 CGTGCATCGCGCGAGACCTGTGACCTGTGTGCTGCGCGAGGCGCGCGCGCGCGCG 1196
Qy 138 AGTCG 194
Db 1197 GTTCCCG 1256
Qy 195 GGGGAAAGCG 254
Db 1257 GGGGAGGACGTACCG 1316
Qy 255 CTCTTGCG 314
Db 1317 CTCTTGCG 1376
Qy 315 GAGAGTGTGCG 374
Db 1377 GAGAGTGTGCG 1436
Qy 375 GAAAAGCGGAGTGTTCATCGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 434
Db 1437 GAGAGTGTGCG 1496
Qy 435 GAGCG 494
Db 1497 GCGGAGCGTGTGCG 1550
Qy 495 GGGCGGAGTGTTCATCGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554
Db 1551 GGGCGGAGTGTTCATCGGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1610
Qy 555 TTGCTGCTGCGTGTGCG 614
Db 1611 CTCTCTGCTGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1670
Qy 615 CAGCG 674
Db 1671 CAGCG 1730
Qy 675 GAAAGCG 734
Db 1731 CCGCGAGGCG 1790
Qy 735 GTTCGAGCGAGCGAGGCG 794
Db 1791 CACGACCTGTGCGAGGCG 1850
Qy 795 GAGCG 854
Db 1851 CAGCG 1910
Qy 855 CAGCG 914
Db 1911 CCG 1970
Qy 915 GAGCG 974
Db 1971 CAGCG 2030
Qy 975 GCTTGTGAGCG 1034
Db 2031 GCTTGTGAGCG 2090

Qy 1035 CGCCAGCG 1094
Db 2091 GGGGAGCG 2150
Qy 1095 GGGGATGATGAGGCG 1154
Db 2151 GGGGCG 2210
Qy 1155 CCG 1214
Db 2211 GCTGCG 2270
Qy 1215 GCG 1274
Db 2271 GACCGAGCG 2330
Qy 1275 GCGGATGAGCG 1334
Db 2331 GCGGATGAGCG 2390
Qy 1335 ACCG 1376
Db 2391 GAGCG 2450
Qy 1377 GCGAAGGACCG 1436
Db 2451 GCG 2510
Qy 1437 CTACCTTTCG 1496
Db 2511 GCG 2570
Qy 1497 GAGCG 1556
Db 2571 GAGCG 2630
Qy 1557 TGCG 1616
Db 2631 CCG 2690
Qy 1617 CAGCG 1676
Db 2691 CAGCG 2750
Qy 1677 GTACGATGATGAGCG 1736
Db 2751 GCG 2810
Qy 1737 GAGCG 1796
Db 2811 GAGCG 2870
Qy 1797 GCTGCTGAGCGAGCG 1856
Db 2871 CCG 2930
Qy 1857 CCG 1916
Db 2931 GCG 2990
Qy 1917 GCTGCG 1976
Db 2991 ACTGACCG 3050
Qy 1977 TGCG 2036
Db 3051 CCG 3110
Qy 2037 GCGCGAGCG 2096
Db 3111 GACCGAGCGAGGCG 3170
Qy 2097 GGTGACG 2156

Db	3171	CGTCAAGCGACCGGAAAGCGGTGTGTCTCTCCGGCGACGAGGCCCGCTCAGCGACTGTGC	3230
QY	2157	GCGCGCGATGCGCGCGCGCGGGGGCGCGAACCGAGCGCTCCAGCTTCGATGTGCTTCCA	2216
Db	3231	GCGGAGTGTGGCCCCCGCGCGCGCGGAGACGAGAGGGCTGTGGGTTCAGCCAGCTTCCA	3290
QY	2217	CTCAACCGCTCATGGCCCCGATGCTGTGGAGGGCGCTTGTGGGGCTGTGTGGCCGAGTCGAGAGCTA	2276
Db	3291	CTCCGCCCACTTGAGACGCACTATGACCGAGGCGTTTCCGAGGTTCGACGAGGGGTGTCTTA	3350
QY	2277	CCGGCGGCGCGTGCATGCTGCTCTGTCAAGCAATCTAGCGCGGAGAGGCTTGCACAGACGAGT	2336
Db	3351	CAGCGCGCGCGTCCCTCCCGGGTGTCTCAGCTCACCGGGCCCCCGTCAACGACGAGCT	3410
QY	2337	GAGCTCGCCGGGCTATTGTGGTGTGCACACGCCCGAGAGGTGTGGGCTTCCGCGATGAGACT	2396
Db	3411	CCGCGAGCGCGGAMACATGTGGTGTGCAGACGTCGGGAGACGGTGTGCTTCCACGACGGGT	3470
QY	2397	GAAGGCGCTGCACCGCGCGGTGTGGGACCTTGTGCAGGTGCAGTCCGAATGACGCT	2456
Db	3471	CCGGCGCTGTGGCGACCGCGGGGCGACCGCTTCTTGAAGGTTCGGGCCCGCGCGGTGTCT	3530
QY	2457	GCTCGGCGTGTGTGCTGTGCATGTCCGCGGACGCCCCGCTC--GGCGTGTCTCGCATGTCTC	2513
Db	3531	GACGCGCGCGCGCAAGCGCGATGCTGTGCCCGGACGCGCCCCCGAGACGTTTGTGCTCCCGTGTCT	3590
QY	2514	GCGCGCTGTGGGAGTGAACGAGCGCGGCGACCGATGCTGTGAGAGCGTGTCCGCGGGAGCTCTGGGCGGT	2573
Db	3591	GCGGCGCGCGACGGCCCCGAACTCCAGTCCGTGTCTACAGGCGCTTCGCGACGGCCACACGAT	3650
QY	2574	CGGTGGCGTGTGTCTGTGGGCGGCGCTTTCCTTCAGGAGGGGCGGCGGATGTCCGCTTGC	2633
Db	3651	CGGCGCTTCTCGCGCGCGGTGTGGGAGCCCGCTGTCTCCCAAGGCCGAGACGCGCGGTGACGTGCC	3710
QY	2634	CACGCTACCTTTGGAGGCGCGAGCGGCTCTGTGATGACACGAAACCG	2680
Db	3711	CACGTAGCGCTTCCAGCGCGGCGACATATGGCTGTGCGCGGCAATGGCCG	3757

```

RESULT 5
US-10-267-255-74
; Sequence 74, Application US/10267255
; Publication No. US2003012469A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600,456US1
; CURRENT APPLICATION NUMBER: US/10/267,255
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 12249
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; US-10-267-255-74

```

Query Match	19.1%	Score 814.2;	DB 14;	Length 12249;
Best Local Similarity	57.6%	Pred. No. 7e-170;		
Matches 1549;	Conservative	0;	Mismatches 1108;	Indels 30; Gaps 4;

Oy	18	CGAGCGGCAAGCCGAATCCGATTCGATTCGTGGAGCAGAGTTGGCGTCTGGCCGGTGG	77
Db	1077	CGAGGACGCGGCCCCCGCAGCCCGCTCGGAGATCTGTGGCATGGCGTGGCCCTTCCCGGGGA	1136
Oy	78	CGTATCGATCTGAGCGGGTTCTTGAACGCTCTTCGAGGGCTCGCGCAACCGTCGGAGC	137
Db	1137	CGTGGCATCCCGGACCACTGTGGCAGCTGTGCGCGAGGGCCGGGACCGCGTCAACGA	1196
Oy	138	AGTCCCGCGCGAAC---GCTGGGATGAGAGAGGTGGTTTGATCCGAGCCCGATGCGCC	194
Db	1197	GTTCCCGCGGACCTGGGGCTTGGAGCTTCGAGCGCGTCTTAGACCCCGAGCCGGACATCC	1256
Oy	195	GGGGAAGACGCCCGTTACGCGCGCATCTTTCCTAGCGAGCTAGCCCTGCTTCAGCGCTC	254
Db	1257	GGGAGAGACGTAAGCGCGCCACCGCGGCTTCTCTAAGAGACGCGCGCGGATTCAGCGCGC	1316
Oy	255	CTTCTTTGGGCACTTCGCTCTCGGAGGGCTTCGGATTTGACCTTCGACATCGACTTGTCT	314
Db	1317	CTTCTTTGGGATCAACGCGCGCGAGGGCTTCGCATAGACCCCGAGAGCGCATGATCAT	1376
Oy	315	GAGGTGTGCTGGAGGCGCTGAGAGACGCGCGATGCGTCCATCGGCGCTCGTCGGTAC	374
Db	1377	GGAATCTCTCGGAGGGCGTTTCAGACAGGGGCGTCTGACGCAACACCTTCGGGGGGA	1436
Oy	375	GGAACGGGAGTGTTCATCGGAGTCGGCCCGTCCGAATATGAGCCCGCGCTCCGCAAC	434
Db	1437	GGAAGTGGGCGTCTTCCTCGGCTTCMAACAGCAACGATCACTGATCAACGTCGACGC	1496
Oy	435	GACGGCGTCCGACAGATTCACGCTCATGGCGGGCTGGGAGCATGCGCCAGCTCGAGAC	494
Db	1497	GGGAGAGCTGCGCGAGGGCTTC-----ATCGGAGCCGGCACTCCGCGACGATCTCTC	1550
Oy	495	GGGCGGAATCTCGTATGCCCTCGGGCTGCAGAGGCGCGTGTGTCGGGTGATACGACTTA	554
Db	1551	CGGCGCGGTGACCTACACTTCGCGCTTCAGAGGGCCGGCGGTGTCGTCGACACGCGCTG	1610
Oy	555	TTGCTCTTCGCTGTGGCCGTTTATCTGGGCTGTGACAGTTTGGGCTCCGGGGAATGCTC	614
Db	1611	CTCTCTCTCGCTGTGTGCGCTGCACCTGGCCGCGCAATCTCTGTGGGAGGGAGTCTC	1670
Oy	615	CACGGCCCTGAGCTGTGGGATTCGATGTTGTTCGCCGAGCACCTCGTGTGACTCTC	674
Db	1671	CTGTGGCCTGGCGGGCGCGCGAGAGTGGCAGCGCACGCGCTTTCATTCAGATTCAAG	1730
Oy	675	GAMAGCCCGGCGCTGGCCAGGAGCGGTGCTGCAGAGCAATTTTGGCGGAGCGGATGG	734
Db	1731	CCGCGACCGGGGCGCTGGCCCCCGACCGGCGCGCTGMAATCTTCTTCGCGCAGACCGCGACGG	1790
Oy	735	GTTTCGACGAGAGGCAAGGATGCGCGCGCTGTGTGTCCTACACCGGCTCAGTGAAGCCCGACG	794
Db	1791	CACCACTGTGTCCGAGGGCGCGCGGTGTGTGTGTGGCCGGTCTTCGGACCGCGCGC	1850
Oy	795	GGAAGCGGATGGAATATTGGGGGTGATTTCAGAGATCCGCAATCAATACGACGGTGGAG	854
Db	1851	CTTGGGCTAACCCCGCTGACGCGGTCAATCCGAGGAGCGCGTCAACAGACCGCGGAG	1910
Oy	855	CAGCGGCTGACCGGTGCGGACCGGAGACTCCCAAGAAATGTTGTGAAAGCGGCGCTGGC	914
Db	1911	CGCGGGCTTGAACCGCGCCCAACGAGACCGGGCGAACGCGGATATCCGCGAGGACATGGC	1970
Oy	915	GAGACGAGGCTGCGCGCGCTTTCGATGGGTTATGTTCGAGGACACACGGCACGAGAC	974
Db	1971	CAAGCGACCGGTGACCGCGCAGCGGTTCGACGCGGTGAGAGCAACGCGACCGGACCCC	2030
Oy	975	GCTTGGTGAACCCATCGAAATCCAGCTCTGAATGCGGTATACGCGCTCGGGCGAGATGT	1034
Db	2031	GCTGGGGGAGCCCGATTCAGAGCCCAAGGCGCTTCCTGCAACCTACGGGGGCGCGCGGAGA	2090
Oy	1035	CGCCACCGCGGTGTGATCGGATCGGTGAGAGACCAACCTTTGGCATCTTGTGATTTGGCTC	1094
Db	2091	GGGAGAGCGCTGTGGGTGGGCTGTGTGAATTCGAACCTTGGCGCACACCAAGTCCGGGCG	2150

DB 2091 GGGGAGGCCGCTGTGGCTGGGCTCGCTGAAGTCGAACTGGGCCACCCAGTCCGGGC 2150

QY	1095	GGGATATCACTGGGCTGCGTGAAGAGTGCATTTGCTTCCTTACAGACCGGGAGAGTATTCGGGCA	1154
Db	2151	CGGCGCGGGCGGCGGTCTATCAAGATGTATGATGCCATTCGGGACCGGGACGTCCCGGCAC	2210
QY	1155	CTTCACGCGCGAGGCGGTGAACCCCGGATCTCATGAGGGGTATCTTCGGCTGACCGTTCAC	1214
Db	2211	GCTGACCTCAACGAGGCCACCCCGGGGTGACGTGTCGCGGTGACGTACCGGTCTCT	2270
QY	1215	GCGCGCCCGGACACCGTGGCCCGGACTGGAAATACGCGCGACGCGCGGGGTGAGCTCGTT	1274
Db	2271	GACCGAGGCGCCAGGACTGGCCGGGACACCGGACAGCGCGCGGTGGCGCGCTCTCGTCTCT	2330
QY	1275	CGGATGAGGGGGGACCAACGGGCGACTGTGTGTAAGAGCGCGCGGGGGGAGAGTGCAC	1334
Db	2331	CGGGGTCAAGCGGACCAACGCGCCATGTGATCTTGAAGGGCCCGCGCGCGGAGAGGACCC	2390
QY	1335	ACCGCG-----GCGCGGAGGAGACCGGACAGAGCTGTGATGCTGTC	1376
Db	2391	GGAGCGCGCGCTGCCGGACGTCTCTCGGACCGCGGGGCCCGCTGCGGTGGGTCTTC	2450
QY	1377	GGCAAGGACCGCGTTCAGCGCTTGATGACAGAGCGCGCGCGCTGCGCGACCATTTGAGAC	1436
Db	2451	CGGCGGACGAGGAGCGCGCGCTCCGAGCGGACGAGCGCGCGCTGGCGCGCCACTGACGCG	2510
QY	1437	CTACCGTTCCGAGTGTCTGCGGCGATGTGGCGTTCAGTGTGGCGACGACGCGGACGCGAT	1496
Db	2511	GCGCGCGCACTTGTGACACCGGCGGACGTGGGACCGCGCTGGCGACACAGCGGAGCGGCTTT	2570
QY	1497	GGAACACCGGCTCGCGGTGCGGGGACGTGAGGGAGGGGCTGCGGGGACGCGCTTGACG	1556
Db	2571	CGACCAACGGGCGCGCGCTGCTCGCGCGGGACCGTGAAGAACTGCTCGCGGCGCTCGCGG	2630
QY	1557	TGCGGCGAGGAGACAGAGTGCGCCCGGTGCGGTGCGCATTCGCGCATTCGACGCGG	1616
Db	2631	CCTGGGCAACGGAAACCGCGCGCGCGCGCTGATCAACGCGCGGACCCGCGTCCGCGG	2690
QY	1617	CAAGCTGCGCTTTCTCTTACACCGGACAGGGGGCGGACGCTGGGCATTTGGCGCTTG	1676
Db	2691	CAAGGCGCGCTTCTCTTCAACCGGACAGGGGACGACGACGCGCGCGCATTTGGCGGAACT	2750
QY	1677	GTACGATGTATGTGTCGCGGCTTCGCGGAGGCGGTTGCACTGTGCGTGAAGGCTGTCAACA	1736
Db	2751	GGCGGCTTCACAGCACCGTGTTCGCGGACGCGCTTGACAGAGTCTGGCGCCACTGACCG	2810
QY	1737	GGAAGCTGACCGGCGCGCTCCGAGAGTGTGTGGCCCGGACCGGCAAGCTGACGCGCG	1796
Db	2811	GCACTTGACCGCGCGCGTGGCGGAGGTCTGTGGCGCGGACCGGACGCGCGGAGGCGCG	2870
QY	1797	GCTGCTGACCAAGACGCTTCACCCAGCGGCGCTGTTCACCTTGCATATTCGCGTCCG	1856
Db	2871	CTTCTCTGACACGAGAGGCTTACCCAGCCCGCGCTGTTCCGCGTGAAGTTCGCGTCT	2930
QY	1857	CGGCGTGTGGGTCTGTGGGGGTGAAGAGCGGAGTGTGCCCGGCGCATACATTCGGA	1916
Db	2931	GCGGCTGTGGAGACTGGGGCTTGGCGCGCGGACATGTGTCCGGGCGCATCGCTGGGGA	2990
QY	1917	GCTGTGGCTGCTGACGCTGTGGCGGCGGTTTCTGCTTACAGACCGCGGTTCTGTGTGC	1976
Db	2991	ACTGACCGCGGCTTACCGCGCGCGGGGTGTGTGCTCGCGGACGCTGCGCGCTGTGTGC	3050
QY	1977	TGCGCGCGGCGCGCTTGATGACAGCGCTGCCGAGCGCGGCGGGCGATGTGTGATCGAGG	2036
Db	3051	CGCGCGGCGCGGCTGACCCAGGACATTCGCCGCGGCGGCGCGCATGTGCGCGTGAAGG	3110
QY	2037	GCGGAGGCGGATGTGTGTGTGGCGGTGGGCGCGGACGCAACGTCGTGTGATTCGCGCG	2096
Db	3111	GACCGAGGACGAGGTGTGCGCGCCCACTGCGCGGACGCGCGCGCGCGGCGGTGACATTCGCG	3170
QY	2097	GGTCAACGCTCCGACACGAGTGTATCGCGGCGCGCGGACACCGGTGACATGCGATCGC	2156
Db	3171	CGTCAACGGAACGGAAAGGTGTGTCTTCGCGGACAGAGCGCGCGTCAAGGACCTGTGC	3230
QY	2157	GCGGCGATGGCGCGCGCGGCGCGGACCAAGCGCTTCAGCTTTCGATTCGCTTCA	2216

Db	3211	GGCGGAGTGGGGCCCGCCCGCGCGGGAGACGAGAGGCTTCGGGCTCAGACGACGCTTCCA	3290
Qy	2217	CTCACCGCTCATGGGCCCCGATGCTGAGAGCGTTTGGGCGTGTGGCCGAGTCGGTGAAGTA	2276
Db	3291	CTCCGCGCCACCTGGAGCCCATGATGACCGAGAGCGTTTCGCGAGAGTCCGACGAGGGGTGTCTTA	3350
Qy	2277	CCGGCGGCGCCGTCGATTCGTCTTCGTCTACCAATCTGAGGGGAGAAAGCTTGCAACAGACGAGT	2338
Db	3351	CAGCGCGCGCTTCCCTCCCGGGTGTCTCCACGCTACCGGGGCCCCGTCACTCGACGAGACT	3410
Qy	2337	GAGCTCGCCGGGCTATTGTGGTGCACCACCGCGCGAGAGATGTGTGCTTCCGCGATGAGT	2396
Db	3411	CCGCGAGGCGCGGAACCTGGGTGCGCGCACGTCCGGGAGAACGTTGCGCTTCCACGACGCGGT	3470
Qy	2397	GAAAGCGCTGCACGCGGCGCGGTGCGGACCTTGTCTGAGGTGCGTTCGAAATGCACGCT	2456
Db	3471	CCGCGCCCTTGGCGACGCGCGGGGCGCACCGGCTTCTGAGAGTGGGCGCGGCGCGTGTCT	3530
Qy	2457	GCTGGGCGTGTGTGCTGCCTGACATGCGCGAGACGCCCGGCGC---GGCGCTCTCGGATGTC	2513
Db	3531	GACGCGCGCGGCAACCGGATGCTGTGCCGAGCGCGCCCGAGACGTTTCGTCCCGTCT	3590
Qy	2514	GCGCGCTGCGGCTGACGAGCGCGGCGACCGTGTCTGAGACGCTCCAGCGGAGCTCTTGAGCGGT	2573
Db	3591	GGCGGCGCGGACGCGGCCGAAACCGAGTCCGTGTACGCGCGGTGCGGACGCGCCACAGAT	3650
Qy	2574	CGGTGGCGCTGTCTCTCTGGGCGCGGCTTTCCTTCAGGGGGGCGGCGGATGCGCGTGC	2633
Db	3651	CGGCGCTTCGCGCGGCGTGGAGCCCTCTGCTGCCCAAGGCCGCGACGCGCGTGGACCTGCC	3710
Qy	2634	CAGCTACCCCTTGGCAGGCGGAGCGCTCTGGATGCACACGAAACCG	2680
Db	3711	CAGGTAGCCCTTCCAGGCGCGGACCTACTGTGCTGCGGAGATGGCCG	3757

```

US-09-953-348-96
RESULT 6
US-09-953-348-96
? Sequence 96, Application US/09953348
? Publication NO. US2003013498A1
? GENERAL INFORMATION:
? APPLICANT: Sherman, David, H
? APPLICANT: Mao, Yinggang
? APPLICANT: Varoglu, Mustafa
? APPLICANT: He, Min
? APPLICANT: Sheldon, Paul
? TITLE OF INVENTION: MITOMYCIN BIOSYNTHETIC GENE CLUSTER
? FILE REFERENCE: 600.530US1
? CURRENT APPLICATION NUMBER: US/09/953.348
? CURRENT FILING DATE: 2001-09-12
? PRIOR APPLICATION NUMBER: PCT/US00/06394
? PRIOR FILING DATE: 2000-03-10
? PRIOR APPLICATION NUMBER: 09/266965
? PRIOR FILING DATE: 1999-03-12
? NUMBER OF SEQ ID NOS: 153
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 96
? LENGTH: 18331
? TYPE: DNA
? ORGANISM: Streptomyces lavendulae
US-09-953-348-96

```

	Query Match	Score	ID.4	DB	Length
18	CGACGGCGACCGCAATCGATTGATTCGTGCGACGAGTTCCTCTCTGCCCCGATG	19.1*	57.6*	1107	1831
1077	CGAGACCGCGGCGCCCGACGCGCCGTCGTGGATGTGGACATGCGCTTCCCGGAGGA	19.1*	57.6*	1107	1831
78	CGTATGATTCATGAGCGGAGTTCTTGACGCGTCTCTGAGGAGCTCGCGACACCGTGGAGC	19.1*	57.6*	1107	1831
1137	CGTGGCATTCGCGGACGACCTGTGGCAGCTGTGTGCGGAGAGGCGCGGACGCGGTACCGA	19.1*	57.6*	1107	1831

OY	138	AGTCCCGCCCGAAC---GCTGGGATGACGACGCGTGGTTGATATCCCGACCCCGATGCCCC	194
Db	1197	GTTCCCGCCGACCGGGGCTGGGAGGTGAGCGCGGTCTACACCCCGACGCGGGCACCCC	1256
OY	195	GGGGAGACGCGCCGTTATGCGCGCATCTTTCTGAGCGACGTAGCTTCTTCAGCGCTC	254
Db	1257	GGGACAGACGTACCGCGGCCACACGGCGGTTCTTCAAGGACCCCGCGGATTCAGCGCCG	1316
OY	255	CTTCTTCGGCATCTTCGCTTCGCGAAGCGCTGCAGATGGAACCTTCGACATCGACTTGTCT	314
Db	1317	CTTCTTCGGCATTCACGCGCGCGCGGCGCTCGCATGGACCCGACGACGCCCATGATCAT	1376
OY	315	GGAGGTGTCGTGGGAGGGGCTGGAGAAAGCGCCGATGCTGCTCATTCGCGGCTCGTCGATAC	374
Db	1377	GGAGGTCTTCGTGGAGGGGCTTCGACAGAGCGGGCTTCAGCGACCAACCTTCGCGGGGGA	1436
OY	375	GGAAACGGAGATGTTTCATCGGGAATCGGCGCGCTCGGAATATGAGGCGCGCTCGCGCAAGC	434
Db	1437	GGAGGTGGGCGGTCTTCGTGGGTCCAAACAGCAAGACGTACCTGATTCACGTCGTGACGC	1496
OY	435	GACGCGCTTCGCGACGATGACGCTCATGCGGGGCTGGGGACGATGCCCACTCGGAC	494
Db	1497	GCGGAGCGTCGCCGAGGGCTTC-----ATGGGACCGGCAACTCCGACACATCTCTC	1550
OY	495	GGGCGGAATCTCGTATGCGCTTCGCGGCTTCGAGAGGCGGTGTTCGCGGGTGAATACGCGCTA	554
Db	1551	CGGCGGCGTCCGCTACACCTTCGCTTCGAGAGGCGCGGCGGTTCGTGACACCGGCTG	1610
OY	555	TTTCGTCTCGCTGTGGCGCTTCATCTGGCTTCGACAGCTTTCGGCTCCGGGGAATGCTC	614
Db	1611	CTCCTCTTCGTGTGTGCGCTGACCTTGCGCGCGCATCTTCGCGGACGGGGGAGTGTCTC	1670
OY	615	CACGGCCCTGGCTGGTGGGGGATATGCGTATGTCGCGGAGCAACCTCGGTGAGCTCTC	674
Db	1671	CTTGGGCTGTGGGCGGGCGGAGCGGTATGAGCCAGCGGACCGCTTCATCGAGTTGAC	1730
OY	675	GAAAGACCCGCGCGCTGCGCCAGGAGCGGTGCGTGCAGAGGACTTTTCGCGGAGGCGGATGG	734
Db	1731	CGGCGAGGGGGGCTGGCCCCCGGACGGCGGCTGCAGAGTCTTCTTCGGGGAACCGCGACGG	1790
OY	735	GTTGCGAGAGAGGCGAAGGCTGCGCTGTGTGTCTTCAAGGCGCTCACTGAGACCCCGC	794
Db	1791	CACCACTGTGTCCAGGGCGCGCGCTGTCTGTGTGGCGCGGCTTCGAGCGCTCGCG	1850
OY	795	GGACGCGCATGGAATATTGGCGGTGATTGAGAGATCCGCGATTCATACGACGCTGCGAG	854
Db	1851	CCTGGGCTTACCCCGTGCACCGGGTATTCGGGGGAGCGCCGCTCAACAGACGGCGGAG	1910
OY	855	CAGCGGTTCACCGTGCAGAACGGGAGCTCCGAGAAATCGTCTGAACCGGCTCTGGC	914
Db	1911	CGCGGGCTTCGACCGGCGCCCAACGAGCCGCGCCACAGCGGGTGTATCCGGGACGACTGGC	1970
OY	915	GGAGCGCAGGCTGCGCGCGGCTGTGGGTGAGGTTATGTGAGGACACCGGCAAGGCGACAGC	974
Db	1971	CACGCAAGGCTGATGGCGCGACAGCGTGAACGCGGTGAGGACACCGGCAACCGGACCCC	2030
OY	975	GCTTGTGACCCCATCGAATTCAGACTCTGAATGCGGTATACGCGCTTCGGGCGAGATGT	1034
Db	2031	GCTGGGACACCGAGTACGAGGCCAAGGCCCTTCGTGGCCACTTACGGGCGGGCGCGGGGA	2090
OY	1035	CGCACACGCGCTGTGATCGGCTCGGTGAAAGCAACTTGGCCATCTGAGTATGGCTC	1094
Db	2091	GGGACGGCGCGTGTGGCTGTGGCTGCTGAATTCGAACTGGGACCAACCCAGTCCGGGC	2150
OY	1095	GGGAGTCATCTGGGCTGTGAAGTGTCTTGTCCCTTACGACGGGCGAGATTCCTGCGCA	1154
Db	2151	CGGCGCGGGCGCGCTCATCAGATGTGTATGGCCATGGCGACCGGAGCGCTCCCGCAC	2210
OY	1155	CCTCCACGCGAGCGCTGAACCCCGGATTCATGGGATGATCTTCGCTGACCGTAC	1214
Db	2211	GCTGACCTTCAGAGGCCACCGCGCGGTGACTGTGCCCGGTGACGTACGAGCTCT	2270

QY	1215	GC	CGGCCCGACACCGGTGGCCCGACGTGAATATAGCCCGACACGGGCGGGGGTGTAGCTGTT	1274
Db	2271	GAC	CGAGGCGCCAGGACCTGGCCCGGACACCGGACAGCCCGCCGTGGCCGCTCTCTGCTCTT	2330
QY	1275	CG	GATGAGCGGGGACCAACCGGACGATGGTGTGTGAAGAGGGCGCGCGCGACAGTGTAC	1334
Db	2331	CG	GCTGACAGCGGACCAACCGCCCATGTGATCTTGTAGAGGCGCCGCGCCGACAGAGGACCC	2390
QY	1335	AC	CGCG-----GGCGCGAGCGACCGGACAGAGCTGTGTGTGTCTC	1376
Db	2391	GAG	CGCGCGCGCGTGGCGGACGTCCTCTGCGACGCGCGGGGCGCGCGTGGCGTGGGTGCGTCTC	2450
QY	1377	GG	CAAGAGACCGCGCTACGCTCTGTGATGCAACAGCGCGCGCGCTGCGCGACCATCTGTAGAC	1436
Db	2451	CG	CGCGCGGACGACGAGCGCGCGCTGTCCAGGCGCAGAGCGCGCTGTGGCGCGCTCACCTGTACCG	2510
QY	1457	CT	ACCGCTTTCGAGTGTGTGGCGCATGTGGGGTTTCAGTCTGTGGCAGACGACGCGAGCGGAT	1496
Db	2511	GC	CGCCCGCACCTTGGACACCGGCGCAGCTGGGCGACCGCGCTGACCAACAGCGGGGCGGCTTT	2570
QY	1497	GG	AGACACACCGGCTCGCGGTGGCGGAGCATGTGAGGAGAGGGGCTGGGACAGCCCTGTGACGC	1556
Db	2571	CG	ACCAACCGGGGCGCCGCTGTCGGGCGGGGAGCCGTGTAGAGAACTGTGTGCGCGGCGCTCGGCGC	2630
QY	1557	TG	CGGCGCAGAGGACAGACGTGCGCCGGTGGGTGGCGCATGTGCGCGATTTCTACGCGG	1616
Db	2631	CT	TGGGACACCGGAACCCGCGCGCCCGGCTGTGTACACGCGCGGAGACCCCGCGCTGTCGGCGG	2690
QY	1617	CA	ACCTCGCGCTTCTCTTTCACCGGACAGAGGGGCGCAGACGCTGTGGACATGGGCGGTGGGCT	1676
Db	2691	CA	AGGCGCGCTTCTCTTTCACCGGACAGGGGAGCAGACAGCCCGGACATGGGCGCGGACCT	2750
QY	1677	GT	AGATGTATGTGTCGCGGTTCCGAGAGGGGTTTCGACTTGTGCGTGTAGGCGTGTTCACCA	1736
Db	2751	GG	CGGCTTCACAGCACCGTGTTCGCGGACCGCTGTGAGACAGAGTCTGCGCCCAAGCTTCACCG	2810
QY	1737	GA	AGCTTCGACCGGCGCGCTCCGAGGTATGTGGGCGGAACCGGACAGGCTTCAGCGCGC	1796
Db	2811	GA	CACTTCGACCGGCGCGCTGGCGGAGGTGTGTTCGCGCGGAGCGCACGCGCCGAGGCGCGC	2870
QY	1797	GC	TCTGTGACCAAGACAGCTTCACCCGACCGGCGGCTTTCACCTGTGAAATATGCGCTCGC	1856
Db	2871	CT	CTCTGTGACACGAGCGGCTTACCCGACCGCGCGCTGTGTGCGCTGTGAGGTGCGGCTGCT	2930
QY	1857	CG	CGCCTGTGGCGGTCGTGGGGGTGTAGAGCGGAGTTGTTGTCGCGGACATATAGCATTCGGTGA	1916
Db	2931	GC	GCTGTGGAGGACCTGGGGCTTGGCGGCCCGCATGTGTGGCGGGCCACTTCGGTCGGGCA	2990
QY	1917	GC	TGTGTGGCTGCTGGCGTGGCGGCGGCGGTCTTGTGTGAGAGACCGGCGTTCCTGTGTGC	1976
Db	2991	ACT	GACGCGCGCGCTACGCGCGCGGGGAGTCTGTGTGCTGTGCGGACGCGCTGTGCGCGCTGTGCGC	3050
QY	1977	TG	CGCGCGGGGCGCTTATGTCAAGGGCGTGGCGCGCGGGGGCCATGTGTTCATTCAGAGC	2036
Db	3051	CG	CGCCCGGCGCGGTGACCAAGGACCTGGCGCGGGGCGGCATGTGTCCGCTGGCAGGC	3110
QY	2037	GC	CGGAGGCGCATGTGGCTGTGCTGTGGCGGCGGACGACGAGCGTGGGTGTTCGATTCGCGCGC	2096
Db	3111	GA	CCGAGGACGAGGTGGCGCGCCCACTTCGCGGACGCGCGCGCGCGGTGTGACATTCGCGCGC	3170
QY	2097	GG	TACACGCTTCGAGACCAAGTGTGATCGCGGCGCGGGCAACCCGTGTATGCGATTCGC	2156
Db	3171	CG	TACAGAGGACCGGAGCGGGTGTGTGTCCGGGGAAGAGGCGCGCGCTTCACCGGACCTGGCGC	3230
QY	2157	GG	CGGCGCATGGCGCGCGCGCGGGGCGGACATCAGGCGCTTCAGAGTCTTCGATGCGTTTCCA	2216
Db	3231	GC	CGGAGTGGGCGCGCGCGCGCGGAGACCAAGGAGCTTCGGGTGTGCGCACCGCTTTCGA	3290
QY	2217	CT	CACACCGCTATGGCGCCCGATGTCTGTGAGGGGCTTGGCGGTGTGGCGAGTCTGGGTGACCTA	2276
Db	3291	CT	CGCGCCCACTTGTAGCGCATGACCGAGGGGTTTCGCGAGGTTCGACAGAGGGGTTCCTA	3350
QY	2277	CG	CGCGCGCGCGTGCATGCTGCTGTGTACACCAATCTGAGCGGGAAGGCTTTCACAGACGAGT	2316

Db 3351 CAGGCGCGCTGCTCCCTCCGAGTGTCTCCACGCTCACCGGGCCCCCTTCACGACGAGCT 3410
Qy 2237 GAGCTCGCGGGGCTATTGGTGGCGCCAGCGCGAGAGGTGGGCTTCGGGAGTGAAGT 2396
Db 3411 CCGAGGCGGGAACCTGGGTGGCGGACGCTCGGGAGACGGTGGCTTCACGACGAGCT 3470
Qy 2397 GAGGCGCTGACGCGCGCGGTGGCGGAGCACTTCGTCAGATCGGTCCGAATCGACGCT 2456
Db 3471 CCGGGCCCTGGCGACGCGGGGGCCACCGCTTCCTGAGGTGGGCGCCGGCGGTGTCT 3530
Qy 2457 GCTGGGCTGTGCTGCTGCTGCTGATGCCGAGCGCCCGGCT--GGCGTGTCTGCACTGCTC 2513
Db 3511 GACGCGCCCGGACCGCATGCTGCTCCGACGCGCCCGCCCGAGAGTTCCTCCGCTCT 3590
Qy 2514 GCGGCGTGGGGGTGACGAGCGCGGAGCGGTGCTCGAGGCGCTCGCGGGGCTTCGGGCGCT 2573
Db 3591 GCGGCGCGCGACGCGCCGGAACCGAGTCCGTGTGACGCGCGCTCGCGAGGCCACACAT 3650
Qy 2574 CCGTGGCTGTGCTCTCTGCGCGCGCTCTCCCTCAGGAGGCGCGCGGTGCTCCGCTCC 2633
Db 3651 CGGCTCTCGCGCGGTGGAGCCGCTGCTGCTCCAGGCGCGGAGCGCGGTGAGCTCTCC 3710
Qy 2634 CACGTACCTTGGGACGCGCGAGCGCTACTGATGACACGAAAGCCG 2680
Db 3711 CACGTACCGCTTCCAGCGCGGCACTACTGCTGCGGGGCACTGGCG 3757

RESULT 7

US-10-267-255-96
; Sequence 96, Application US/10267255
; Publication No. US20030124689A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; TITLE OF INVENTION: Mltomycin biosynthetic gene cluster
; FILE REFERENCE: 600 456U31
; CURRENT APPLICATION NUMBER: US/10/267,255
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US 09/266,965
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: US 08/624,447
; PRIOR FILING DATE: 1996-08-19
; PRIOR APPLICATION NUMBER: PCT/US94/11279
; PRIOR FILING DATE: 1994-10-06
; PRIOR APPLICATION NUMBER: US 08/133,963
; PRIOR FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 96
; LENGTH: 18331
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-10-267-255-96

Query Match 19.1%; Score 814.2; DB 14; Length 18331;
Best Local Similarity 57.6%; Pred. No. 6; 8e-170;
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

Qy 18 CGAGGCGCAGCGGAGATCGATGCGATGCGGAGCGAGTTGCGGTCGCGCGGTGG 77
Db 1077 CGAGGACCGGCGCGGAGCCGTCGATGCTGGGCAATGGCGTCCGCTTCGCGGGGA 1136
Qy 78 CGTATCATCTGACGCGGATCTCTGACGCTCTCGAGGAGCTTCGCGACACCGTGGGCG 137
Db 1137 CGTGCACTGCGCGAGCACTGTGCACTGTGTGCGCGAGGCGCGGAGCGCGTACCGA 1196
Qy 138 AGTCCCGCGGAGC--GCTGGAGTACAGCAGCGGTGTTGATCCGACCCCGATGCC 194
Db 1197 GTTCCCGCGGAGCGGGGCTGGGAGCTGACGCGCTTACGACCCCGAGCGGGGACCC 1256

Qy 195 GGGAGAGCGCCCGTATCGCGCGATCTTCTTGAAGGACGTACGCTGTGACGCTC 254
Db 1257 GGGAGAGCGTACCGCGGCGCACCGCGGCTTCTTGAAGGACCGCGGATTCGACGCGC 1316
Qy 255 CTTTTCGCACTCTGCTGCGGAGCGCTGCGGATGAGCCCTGCACTGCACTTCTTCT 314
Db 1217 CTTTTCGCACTCTGCTGCGGAGCGCTGCGGATGAGCCCTGCACTGCACTTCTTCT 1376
Qy 315 GGAAGTGTGCTGGAGGCGCTGAGGAAAGCGCGGATGCTCATCGGCGCTGTGCTGAC 374
Db 1377 GGAAGTGTGCTGGAGGCGCTGAGGAAAGCGCGGATGCTCATCGGCGCTGTGCTGAC 1436
Qy 375 GGAAGCGGAGTGTTCATTCGCGATCGCGCGCTCGGAATATAGGCGCGCTGCGCAAGC 434
Db 1437 GGAAGTGTGCTGGAGGCGCTGAGGAAAGCGCGGATGCTCATCGGCGCTGTGCTGAC 1496
Qy 435 GACGCGCTCGCAAGATCGACGCTCATGCGGCGCTGGGAGCATGCCAGCTGAGC 494
Db 1497 GCGGAGCGTCCCGAGGCGCTTC-----ATCGGAGCGGCAACTCCGCACTCTCTC 1550
Qy 495 GGGCGGATCGTATGCGCTCGGCGTGGAGCGCGGCTGCGGCGGATGATAGGCTTA 554
Db 1551 CCGCGGCTGCTTACCTTTCGCTTTCGAGGCGCGGCTGCTGCTGCTGCTGCTGCTG 1610
Qy 555 TTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 614
Db 1611 CTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1670
Qy 615 CACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674
Db 1671 CTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1730
Qy 675 GAAAGCGCGGCGCTGCGCGAGCGGATGCTGCAAGGCACTTCGCGGAGGCGGATG 734
Db 1731 CCGCGAGGCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1790
Qy 735 GTTTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 794
Db 1791 CACCACTGCTGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1850
Qy 795 GAGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 854
Db 1851 CTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1910
Qy 855 CAGGCTGTGACCGCTGCGGAGCGGAGCTCCCAAGAAATGCTGCGAAACGCGGCTGCG 914
Db 1911 CCGGCGCTGACCG 1970
Qy 915 GAGCGAGGCTGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974
Db 1971 CAAAGCAAGCTGACG 2030
Qy 975 GCTTGTGACCCCATTCGAATTCGAAGCTTGAATGCGGTATACGCGCTCGGCGAGATGT 1034
Db 2031 GCTGAGCGACCCGATTCAGAGCGCGCGCGCTCTGCGCACTACGCGCGCGCGCGCG 2090
Qy 1035 CGCAAGCGCGCTGCTGATTCGCGTGGTGAAGCAACACTTGGCGCATCTGATGATGCTC 1094
Db 2091 GGGAGGCGCGTGTGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2150
Qy 1095 GGGGATCACTGGGCTGCTGAGGATGCTGCTTGTCTTTCAGACGCGGAGATTCCTGCGCA 1154
Db 2151 CCGCGCGCGGCGGCTCATCAAGATGATGATGATGATGATGATGATGATGATGATGATG 2210
Qy 1155 CTTTCAGCGCGAGCGCTGAACCCCGGATCTCATGCGGATCTTTCGCTGACCGCTAC 1214
Db 2211 GCTGCACTTCAAGAGCCACCCCGCGCTGATGATGATGATGATGATGATGATGATGAT 2270
Qy 1215 GCGGCGCGGAGCACGCTGCGGAGCTGGAATACCGCGCGAGCGCGGCGGCTGATGCT 1274
Db 2271 GACCGAGGCGGAGATGCGCGGAGCACCGGACCGGAGCGCGCGCTGCTGCTGCTT 2330

QY 1275 CGGATGAGCGGAGCAACGCGCAGTGTGCTGAGAGGAGCGCGCGCGAGCGTGCAC 1334
DB 2331 CGGCGTCAAGCGGACCAACGCCATGTGATCTTGAAGGCGCGCCCGCCGAGAGGACAC 2390
QY 1335 ACCGCG-----GCGCGGAGCGACCGGCGAGAGCTGCTGTGCTGTC 1376
DB 2391 GGAACGCGCGCGTGCAGGAGCTGCTCGAGCGCGCGCGCGCGCGCTGCGGTGTGTC 2450
QY 1377 GCGAAGAGACCGGCTGAGCCCTGAGATGACAGCGCGCGCGCTGCGGACATCTGAGAC 1436
DB 2451 CGGCGCGAGCGAGCGCGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGAGCG 2510
QY 1437 CTACCTTCGAGTGTGCTGAGGAGTGTGCGCTTCACTGTGAGGAGCGAGCGCGAGT 1496
DB 2511 GCGCGCGAGCTGAGCTT 2570
QY 1497 GAGAGACCGGCTGCGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1556
DB 2571 CGAGCAGCG 2630
QY 1557 TCGCGCGAGGAGCAGACGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1616
DB 2631 CCGGCGCAGCGGAGCG 2690
QY 1617 CAAGCTCGCTTCTTCAACCGGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1676
DB 2691 CAGGCGCGCTTCTTCTTCAACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2750
QY 1677 GTACGATGTATGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1736
DB 2751 GCGCGCTCAGACGACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2810
QY 1737 GAGAGTTCAGCGCGCGCGCTGCGGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCG 1796
DB 2811 GCACTTCAGCGCGCGCGCTGCGGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCG 2870
QY 1797 GCTGCTGACGACGAGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1856
DB 2871 CCGTGTGACGACGAGCGCGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2930
QY 1857 GCGCGTGTGCGCGCTGCGGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1916
DB 2931 GCGCGTGTGCGCGCGCTGCGGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2990
QY 1917 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1976
DB 2991 ACTGACCGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3050
QY 1977 TCGCGCGCGCGCGCTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2036
DB 3051 CGCGCGCGCGCGCGCTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3110
QY 2037 GCGGAGCGCGATGTGCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2096
DB 3111 GACCGAGGAGCGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3170
QY 2097 GGTGACCGCTGCGGAGCGAGTGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2156
DB 3171 GGTGACCGAGCGGAGCGAGTGTGCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCG 3230
QY 2157 GCGCGCGATGCG 2216
DB 3231 GCGCGAGTGTGCG 3290
QY 2217 CTGACCGCTGATGCT 2276
DB 3291 CTGCGCGCAGCTGAGCT 3350
QY 2277 CCGCGCGCGCGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2336
DB 3351 CAGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3410
QY 2337 GAGCTGCGCGCGCTATGAGGCTGCGCGCGCGCGCGAGGAGTGTGCGCTTGCAGATGAGT 2396

DB 3411 CCGAGCGCGGAGCACTGTGTGCGGCACTGCGCGGAGAGCGTGTGCTTCAAGCGCGT 3470
QY 2397 GAGGCGCTGACGCGCGCGGCTGCGGAGCACTTGTGCGAGTGTGCTGCGAATGACGCT 2456
DB 3471 CCGCGCGCTGCGGAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3530
QY 2457 GCTGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2513
DB 3531 GACGCGCGCGGAGCG 3590
QY 2514 GCGCGCTGCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2573
DB 3591 GCGGCGCGGAGCT 3650
QY 2574 CCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2633
DB 3651 CCGCGCTGCG 3710
QY 2634 CAGCTACCTTGGAGCG 2680
DB 3711 CAGCTACCGCTTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3757

RESULT 8
US-10-205-032-19
; Sequence 19, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianhu
; APPLICANT: Stafla, Alfredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARAMICIN
; FILE REFERENCE: 3016-205
; CURRENT APPLICATION NUMBER: US/10/205, 032
; CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 19
; LENGTH: 5355
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-19

Query Match 18.1%; Score 770.8; DB 14; Length 5355;
Best Local Similarity 54.3%; Pred. No. 2.6e-160;
Matches 1868; Conservative 0; Mismatches 1517; Indels 57; Gaps 13;

QY 32 AAGATCCGATGCGATGCGAGGAGTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 91
DB 101 ACGGCGCGATGCG 160
QY 92 GCGGCTGTGAGCGCTCTCGAGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151
DB 161 AGGACCTGTGCG 220
QY 152 ---GCTGAGAGCGAGCG 208
DB 221 GTGCTGTGAGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 280
QY 209 TTAGCGCGGATCTTCTTCTGAGCGAGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 268
DB 281 GCGGAGAGGCGGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 340
QY 269 GCGCTGTGAGAGCGCTGCGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 328
DB 341 CCGCGCAGAGCGCGTGTGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 400
QY 329 AGGCGCTGAGAGAGCG 388
DB 401 AGGCGCTGAGAGCG 460
QY 389 TCATCGGAGTGTGCG 448

Db 461 AGGTGGGTCGCGCCACCAAGGGGTACGGGTGACGCCGTGACGGTTCGGAGGGCGCG 520
OY 449 AGATGACGCTCATGCGCGGCTGCGGAGCAATGCCAGCGTCGAGCGGGCCGAATCTCGT 508
Db 521 AG-----GGGTCCGGCTGACCGGAGCGCGGACGCCGTCTGTTCGGAAGCAATCTCGT 574
OY 509 ATGCCCTCGGGCTGCGAGGGCCGTGTGTGCGCGGTGATACCGGCTTAATTCGTCTCGTGG 568
Db 575 ACCTGCTCGGGCTGAGGGATCCGGCCCTGACCGGTGAGACCGGCTGTGTCTCTCGTGG 634
OY 569 TGGCGCTTCACTGCGCTCTGACAGCTTGCCTCCGGGGAATGCTTCACGGGCTTCGCTG 628
Db 635 TGGCGGTACACTTCGGGTGACAGGCGCTGCGCGGGGCGAGTGCAGGCTGACCTGGCGG 694
OY 629 GTGGGGTATCGCTGATGTTGTGCGGAGACCCCTGTGTGGCTTCGAAGACCGGGGGCG 688
Db 695 GCGGGGTGCGCTGATGCCACCGCGGGGCAATTCGTGAGATTCCTCCGGCAGCGGGGCG 754
OY 689 TGGCCAGGAGCGGTGCTGCAAGGCAATTCGCGCGAGGCGCGATGGGTTCGAGCAGAGCG 748
Db 755 TCGCGCGAGAGCGGCGCTGCGGGGGCTTCGGGGCGGGCGGACCGGCTGGGGCG 814
OY 749 AAGGTCGCGCTCGTGTCTCAAGCGGCTGATGAGAGCCGCGGAGCGGCGATCGGA 808
Db 815 AAGGCGTGTGTGTGTGCTGCTGCAACGGGCTCTCCGACGCGGTGCGGACCGGCTGGG 874
OY 809 TATTGGCGGTGATTCGAGGATTCGCGATCAATCAAGAGGTGCGAGCGGCTTCGACCG 868
Db 875 TGCTGGGCGTATCCGGGGTTCGGGCTTCACCAAGAGCGGGCGAGCAACGGGCTGACCG 934
OY 869 TGGCCGAAGGAGCTGCCAAGAAATGCTGCTGAACCGGGCCCTGCGGAGCGAGGCTCG 928
Db 935 CCCGAGGCGGCCCGCCAGAGCGGGGTATCCGGGCAAGCGGTGACCGACCGCCGCTCG 994
OY 929 CCGGCTCTTCGGTGGTTATGTGAGGCAACAGGCAACGGGCAACGAGCTTGGTGAACCCA 988
Db 995 GCGCGGACCAATTCACGCGGTGAGGCGCAACGGGCAACCGGCTGCGGACCGGCA 1054
OY 989 TCGAAATCCAGACTTCGATGCGGTATACGCGCTGCGGCGAGATGTCCGCAACGCGGCTGC 1048
Db 1055 TCGAGGCGGACGCTGATTCGCGGCTTACGGGCGGACCGGACCGCGGACCGGCGCTCT 1114
OY 1049 TGATCGGCTGCGTGAAGACCAACTTGGCACTTCGATGATGCGTCCGGGATCACTGAGC 1108
Db 1115 GCGTGGCTCGTGAAGTGAACATCGGGCAACGCCAAGCGGCGCGGCGTCCGGCGGCG 1174
OY 1109 TGCTGAGGTGCTTGTTCCTTACAGCAACGGGCAAGTTCCTGCGCACTTCCAACGCGCAG 1168
Db 1175 TGATCAAGATGCTCTGGCGATGCGGGCGGGGACGCTCCACCAACCTGCGACGCGCAG 1234
OY 1169 CGCTGAACCCCGGATTCATAGGGGTGATCTTCGCGTACCGGTACAGCGGCGCGCGGACAC 1228
Db 1235 TCCGACCCCGCTGCTGATGCTGCGCGGTGCTCCGGTGTGACACCGGAGTGTGC 1294
OY 1229 CGTGGCGGAGCTGG---AATAGCGCGGACCGGCGGGGTGAGCTGTTTCGCGATGAGCG 1285
Db 1255 CTTGGCCGCGGTGCGCGGGGCGCGGCGGAGATCTCCGCGTTTCGGGGTGAACG 1354
OY 1286 GGAACCAACGCGACGTGTGCTGAAGAGCGCGCGGCGGACGTTGCAACCGCGGCGGCG 1345
Db 1355 GACACCAACGCGACGTGTGCTGAAGAGCGCGCGGCGGATCCCGGTCCGGTCAACGACCGGCGC 1414
OY 1346 C-----GAGGGAACCGGAGAGCTGTGTGTCTGTCCGCAAGACCGGCTACGCC 1396
Db 1415 CACCCACGAGGACCTTGCCTGTGTGTGTGTGTGTCTTCCGCGGAGCGAGGAGGCGC 1474
OY 1397 TGAATGCAACAGCGCGCGCTGCGGACCAATCTGAGACCTTACCTTCGAGTGTG 1456
Db 1475 TGGCGGCCAGGCGGACCGGTTGCTGACGACCTTGGCGGCGGACCCGACCGGACCGC 1534
OY 1457 GCGATGTGCGCTTCACTGTGCGGACGAGCGGCGGATGAGAGACCGGCTCGGAGTGG 1516

Db 1515 TGAAGTGGATTCCTCTGCGCACAGCCGGGCGCGCTGAGACACCGGCGCTGTG 1594
OY 1517 CGGCAACGTCGAGGAGAGGAGCTGCGGCGACCTTGACGCTGCGGCGGAGGACAGAGT 1576
Db 1595 TGGCGCGGACCGGACCGGCTGCTCGGCTCTGTGACGCGCTGCGCGGCGGACCGG 1654
OY 1577 GCGCGGTGCGGTGCGGATTCGCGGATTCCTGACCGGCGAGAGCTGCTTCTCTTCA 1636
Db 1655 CGGCGGCGCTGATCCGGGCGACGAGTACGATTCGCG---GACCGGTTCTCTTCG 1711
OY 1637 CCGGACAGGGGCGGACGCTGCGGACAGGCGGCTGCGGCTTGAAGTATGATGTCGCGT 1696
Db 1712 CCGGACAGGGGCGGACGCTGCGGATGAGGCGGACGCTGAGAGGCGGTTCCCGCT 1771
OY 1697 TCCGAGAGCGCTTGAACCTGTGCGTGAAGCTGTTCAACAGAGACTCGACCGGCGCTC 1756
Db 1772 TCGCGGAGCGCTTGAACGCTTCGCGCGCGGCTGAGACCGGACCTGATGTTCGCTGC 1831
OY 1757 GCGAGTGAATGTGGGCGGACCGGCGGACGCTGACGCGCGGCTGTCGACGACAGAGCT 1816
Db 1832 TCGCGGTGCTGTTCGCGGAGCGGAGTTCGAGACGCGCGGCTGTCGACGCGACCGG 1891
OY 1817 TCACCAACCGGCGCTTCACTTCGAAATGCGCTGCGCGCTGTGTGCGGCTGTGG 1876
Db 1892 ACGGCAAGCGCGCTTTCGCGGCTGAGGCTGTGTTCCGACTGTGTCAACGCGCTGG 1951
OY 1877 GTTAGAGCGGAGTGTGCGCGGCAACAGCTGAGAGCTGAGTGTGCTGCGTGG 1936
Db 1952 GGTTCGCGGCGGACGCTGTCTGCGGCTGCGGCTGAGGAGCTGAGCGGCGCGGACGTTG 2011
OY 1937 CGGCGGTGTTCTCGCTTGAAGACCGGCTGTTCTGTGTGCTGCGCGGCGCTGATGC 1996
Db 2012 CCGGCTGATTCGAGTGAAGACGCGGTGCGAGCTGTGTGCGGCGGCGGCGGCTGATGG 2071
OY 1997 AGCGCTGCGCGCGCGGCGGAGTGTGTGATGAGAGCGCGGAGCGGACGATGTGCTG 2056
Db 2072 GGAAGCTGCGCTCGGCGGCGGAGTGTGTGCGGCTGCGGCGGACGAGAGGAGTGGGCG 2131
OY 2057 CTGCGGTGCGCGCGGACGCTGCGGCTGAGTGTGCGCGGCTGACGCTTCGAGCACGAG 2116
Db 2132 CCTGTCTG-----ACGGGAGCGGCTGCGGTGCGGCGGCTGACGCGGCGCTGCG 2185
OY 2117 TGCTCATGCGGAGCGCGGCGGACCGGCTGATGCGGAGCGGAGTGGCGCGCGCG 2176
Db 2186 TGTGTGCTCGGCGGAGGAGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCGGCTGCG 2245
OY 2177 GGGCGGAAACCAAGCGCTCAAGCTTCGCACTGCGCTTCACTACCGCTCATGCGCGG 2236
Db 2246 GCAACCGGACCGGACGCTTCAAGCTGAGCCAGCGGTTCCACTCCCGGACGTTGAGGCG 2305
OY 2237 TGCTGAGGCGGTTGCGGCTGTGCGGCGGAGTGTGAGCTACCGGCGGCTGTGATTCGTC 2296
Db 2306 TGTGAGGCGGTTGCGGCGGCTGTGCGGCGGCTGAGCTGATCAATGCCCGGAGATCCCG 2365
OY 2297 TGTGACGAATCTGA---GCGGAGAGCTTGCAGACGAGAGTGAAGCTTCGCGGCTAAT 2353
Db 2366 TGTGTGGAACGAGCGGCGGCTGCGGACCGGCGGAGAGCGGCTGCGGACCGGCTAAT 2425
OY 2354 GGTGTGCGACGCGCGGAGGTGTGCGCTTTCGAGTGAAGAGCGCTGACGCGG 2413
Db 2426 GGTGTGCGGACGTCGCGGAGCGGCTTTCGAGGAGGAGTTCGCGGCGGCGGCGGAC 2485
OY 2414 CCGGTGCGGAGCACTTGTGAGGCTGAGTGTGCGGAGTGAAGAGCGGCTGAGCGG 2473
Db 2486 AGGGAGGCAACGCTTGTGCGGCTGAGGCTGAGGAGGCTGTGCGGCTTTCGAGG 2545
OY 2474 CTTGCT--GCGGAGCGCGGCGGCGGCTGCTGAGCTGCGGCGGCTGAGGCGG 2530
Db 2546 AGTGTCTGCGGCGGACGCGGAGAGTGTGCTGCTGCGGCTGCGGCGGCGGCGG 2605
OY 2531 AGCGGAGCGGCTGAGGCGGCTGCGGCGGCTTTCGAGGCGGCTGAGGCGGCTGCT 2590
Db 2606 AGCCCGGACCTGCTGCGGCGGCTGCGGCGGCTGAGGCGGCGGCGGAGTGAAGT 2665

2591 GGGCCGCGCTCTTCC---CTAGAGGGGGGGGGTGGCCGCTGCCCACTTACCTTGGC 2647
Db GTTCCCGGGGTGTTCCGGGGCACCGGGCGGCAAGGTTCAGCTGCCCACTGACGCTTCC 2725
QY AGCGGAGCGCTACTGATCGACAGAAAGCCGACGCGCGGTGGCGACCGCGCTG 2707
2726 AGCACCGGGGCTACTGGCTGGCGGGGGCCCGCTGGCCCGCGACAGACGCTG-- 2783
QY CTCGGGAGCGGGTCAAGACGAGGTGAGAGGGGGGGCGGTGGCGGGCGACCGGC 2767
2784 --GCGGTACCGGATCGCTGGCGGGCCCTGCGGACCGGTGCGCGCGCGCGACCGAG 2841
QY GAGAGGCTGGCTGACCACTCCGCGCGCGGAGAGCGGAGCGCGGAGAAAGTGG--AGGC 2826
2842 AGGTTGGCGGGGGCGGTGGTGTGGTGTGCCCGCGCAAGAGGGGTGCGCTGCGC 2901
QY GCGGCGGACCGCTTCCGCTCGAGATGATGAGCCAGGGGTGCTTATCACTCTGG 2886
2902 GACGCGCGCGAGCGGGCGCTGACCGGGGGGGGGCCAGGTACCGGGGTGACGGTGAAC 2961
QY CTTGGGGTACGAGCGGGCGCGCGCGCTGGGAGGTGAGATGCGCGTGAACGCG 2946
2962 GCGCGCGAGTGGACCGGGACACCTCGCGCGCGCTGATGACCGAGGGCGCGCGACGCG 3021
QY GCGGGGCTGAGCTTCAATGATGTCAGCTGCGCGCTGGGAGTGGAGATGCGCGTGAACGCG 3006
3022 GACGCG-----GAGCGGAGCGGGGTCTGTCTGTGGGCTTCCAGCAACGCG 3069
QY GGAAGCGCCACCTCTCGCTGCTGCTGCTGAGAGCGAGTGGCGCGGGCGACCTGCGCGCTG 3066
3070 GCACTATCCCGGTCCGCTCGGTGCGCGCGGGGTGCTGGCACTGTGCTTCCCGCAG 3129
QY GCGAGGGCGGTGAACGCGCTGCTGCTGGGCGCAACCGGTGATGCTTTCGGCGGAGCG 3126
3130 GCGCTGACGACCTGGGGGCGCTCGCGCGCGCTGGTGGTGAACCGGGGGGGCGGTGCGC 3189
QY TTTGTACCCAGCTGACCACTGCGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 3186
3190 GTAGAGCGCGGGGAGTCCCGTGGGTGCGCGGAGCGCACTGTGGGGCTTTCGCGCGCGCTG 3249
QY GCGATCGAGCGCGCGCGCACTGCGCTGCGGTGCTGACCGGCACTGATGACCGCTGACAGA 3246
3250 GCGCGGCTGCACTCCCGCGGTCTGGGGCGGGCTGGTGAACCTGCGGTGACCGCGAC 3309
QY ATAGCCCGCTTACGCGCGGGGAGCGGGGTGCTGATGATGCGGCAACCGCGGGGTGCGT 3306
3310 GACCGGGACTGGGACCTGCTGGCGCGCGCTGCGCGCGCGGAGGACCAAGGTGCGCGTC 3369
QY CTCGCGCGGGTGCATGGGGCGGAGCACTGGAGCGCGAGTGCATGCGACCGCGCGGACG 3366
3370 CGGGGGGGGTGGGTGACCGCGCGGGGTGCTGCGCG--GCGCGCGCGCGCGGGCGAGCG 3428
QY CCCGAGAAACGCGCTTACTGAGTGCCTGGCGCGGTGGCGGTATGAGCATTCGCCCTGCG 3426
3429 GCGCTGGGGTCCGCGCGGACGCTGCTGTAACCGGGTACGGGCGCGCTGGGGCGCA 3488
QY 3427 GACCGGTTGCTGCGCGACGTCG 3448
Db 3489 CACGCGCGCGCTGGCTGGCGCGC 3510

RESULT 9
US-10-205-032-1
; Sequence 1, Application US/10205032
; Publication No. US20030113874A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianhu
; APPLICANT: Scatita, Alredo
; TITLE OF INVENTION: GENES AND PROTEINS FOR THE BIOSYNTHESIS OF ROSARMICIN
; FILE REFERENCE: 3016-205
; CURRENT APPLICATION NUMBER: US/10/205,032

CURRENT FILING DATE: 2002-07-26
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 60196
; TYPE: DNA
; ORGANISM: micromonospora carbonacea subspecies aurantiaca
US-10-205-032-1
Query Match 18.1%; Score 770.8; DB 14; Length 60196;
Best Local Similarity .543%; Pred. No. 2.1e-160;
Matches,1868; Conservative 0; Mismatches 1517; Indels 57; Gaps 13;
QY 32 AAGATCCGATTCGATTCGTCGAGGAGTTCGCTGCGCGCGGTGGCGGTGATTCATCGA 91
Db 41149 ACGGCGCATATGCGCTGCTGGAGCATGAGCTCCGCTACCCGGGGGTGCTCTCCGG 41208
QY 92 GCGGGTTCGAGCGCTCTCGAGGGCTGCGCGCACCGGTGCGGAGTCCCGCGGAC 151
Db 41209 AGGACTGTGGCGGTGCTGACAGCGGACAGAGCGATGCGGGCTTCCCGACGAC 41268
QY 152 ---GCTGGATGACAGACGCTGCTTGTATCCCGACCCCGATGCCCGGGAGAACGCCG 208
Db 41269 GTGGCTGGAGCTGCGACCGCTGTACGACCCGGAATCTGATGCGGGAAACACCTACT 41328
QY TTAAGCGGCACTTTCTGAGCGACGACGCTGCTTCCGATGCGCGCTCTTCCGATCT 268
Db 41329 GCGGAGAGGGGGTCTTGGCGCGGAGAGACTTTCAGCGCGCGCTTCTTCCGGGTGT 41388
QY 269 GCGCTGCGGAGCGCTGCGAGATGACCTGTCACATTCGCTGAGAGGTGCTGGG 328
Db 41389 GCGCGACAGAGCGGTGATGAGACCCGACAGCGGCTCTTGTGAGGTGCTTGGG 41448
QY 329 AGGCGCTGAGAGCGCGCGATTCCTTCATGCGCGCTGCTGATCGGAAACGGAGTGT 388
Db 41449 AGGCGCTGAGAGCGGTGCTGCGGACCGACCGCAACAGCTGCGGGCTGCGGACCGGGGTCT 41508
QY TCAATCGGATAGGCGCGCTGCAATATGAGCGCGGCTGCGCGGACGAGCGGCTGCGGAG 448
Db 41509 ACGTGGTGGCGCGCCACAGGGGTGACGGGTGACGCGGTGACAGTGGCGGAGGGCGGG 41568
QY AGATGACGCTCATGAGCGGGCTGGGAGAGATGCCAGCGTGGAGCGGGCGGAATCTGCT 508
Db 41569 AG-----GGGTTCGCGTGAACCGGACGCGCGAGCGCGCTGCTGCGGAGATCTGCT 41622
QY ATGCGCTGCGGCTGCGAGAGCGGTGTGTGCGGTGATATGCGGCTATTCGCTGCTGCTG 568
Db 41623 ACTGCTGGGCTGGAAGGATCCGCGCTGACCGCTGAGACCGGCTGCTGCTGCTGCTG 41682
QY TGGCGCTTCACTGCGCGCTGTCAGAGCTTGGCGCTCGGGGATATGCTCAACCGCGCGCTG 628
Db 41683 TGGCGGTGCACTTGGCGGTGACGCGCTGCGCGGGGAGATGGGCTGGCACTGGCG 41742
QY GTGGGTATTCGCTGATGTTGTGCGGAGACCTCGTGTGCTGCTGCAAGAACCGCGGCGC 688
Db 41743 GCGGGTGGCGGTATGCGGACCGGGGCGGATTCGTGATTCCTCCCGGAGGGGGCC 41802
QY TGGCGAGGAGCGGTGCTGCAAGGCACTTTCGGGAGAGCGGATGGGTTCGAGAGAGCGG 748
Db 41803 TCGCGGCGAGCGGCGCTGCGCGGGGCTTTCGGGGGGGGCGGAGCGCACCGGCTGGGGG 41862
QY AAGGTTGCGGCTGCTGCTCAAGCGGCTCAGTGAAGCGCGCGGAGCGGATCGGA 808
Db 41863 AGGGGTGCTGTGTGCTGCTGCAACGCGCTTCTCGACGCGGTGCGGACGCGCGCTGGG 41922
QY TATTGCGGTGATTCAGAGATCGCGATCAATCAAGAGTGGAGAGCGGTGCTGACG 868
Db 41923 TGTCTGGGTGATTCGGGGTTCGGCGGTCAACAGAGCGGGCGAGCAACGGGCTGACCG 41982
QY TGGCGAGGAGAGCTCCCAAGAAATGCTGCTGAAACGGGCGCTGCGGAGCGGAGCTGCG 928
Db 41983 CCGGAGCGGCGCGCGCGGACGAGCGGTGATTCGCGGAGGCGCTGACGACCGCGGCTGCG 42042

QY	929	CCGGCTCTCGGTGGGTTATGTGAGGCACAACGACACGGGACACAGACCTTTGGTACCCCA	988
Db	42043	GCGCCGACCAAGATCGACCGGTGAGAGCGCACACGAGCACCCGCTCGGACCCGA	42102
QY	989	TGCAATTCGAAGCTGTAATGCGGTATAGCGGCTTCGGGCGAGATGTCCGACAGCGCGCTGC	1048
Db	42103	TGAGGCGCAGGCGCTGATGCGCCCTTACGGGCGCCACCGAACCCCGAACCGGCGCTCT	42162
QY	1049	TGATCGGGTCGAGTAAAGACCAACCTTTGGCCATCCCTGAGTATGCGTCGGGAAATCATCTGGGC	1108
Db	42163	GCGCTCGGCTGTTTAATTCGAACATTCGGGACACCCGACGGGCGGGCGGGGCGGTTCGCGGCGC	42222
QY	1109	TGCTGAAGTGTCTTGTCCCTTTCAGCACGGGCGACGAATTCCTGCGCACTTCACGCGCAG	1168
Db	42223	TGATCAGAATGTCTCTGGCATGTGGGCGCGGAGCGCTCCACCACTTCAGCGCGCAG	42282
QY	1169	CGCTGAACCCCGGATCTCAATGGGGTATATTTGGGCTGACCGTCAAGCGCGCGCCCGAACAC	1228
Db	42283	TCCGAGCCCGCTGTGTGACTGTGCTGCGCGGCTGTCTCGGCTGTGACACCGAGGTGTGC	42342
QY	1229	CGTGGCGGACATG---AATAGCGCGGACGGGCGGGGAGTGGACTGTTTCGGCATAGAGG	1285
Db	42343	CTTGCGCGGTTCGCGGGGCGCGCGGACGGGCGGGAAATCTCGGCTTCGGGGTATAGG	42402
QY	1286	GGACCAACCGCGCATGTGTGTGGAAGAGCGCGCGCGGACGTGACACACCGCGGCGC	1345
Db	42403	GCACCAACGCGCACGTATGTGTGACGACGCGCGCGCGCGGTGTGACGACCGGCGC	42462
QY	1346	C-----GAGCGACCGGCGAGAGGTGTGTGTCTGTTCGCGAAGAACCGGCTCAGCC	1396
Db	42463	CACCCACAGAGACCTGTCCGCTGTGTGCGGTGTCTTCGCGCGGACGAGAGCGGCGC	42522
QY	1397	TGATGTGACACAGCGCGGCGGCTGGCGACCATCTGAGACCTTTCGACGTGTCTTG	1456
Db	42523	TGCGCGCCGACAGCGGACCGGTGTGTATCGGACCTTGCGGCGGACCCCGACGCGACCCGC	42582
QY	1457	GCGATGTGCGCTTCAGTCTGTGCGCACGACGCGCAGCGGATGTAGACACCGGCTTCGCGGTG	1516
Db	42583	TGAGCTGTGGATTTCTCCCTGGCCACACGACCGGGCGCGCTGGAGCACCGGGCGGTGTGG	42642
QY	1517	CGGCGACGTGAGGAGAGGGGCTGCGGGACGCTTGACCTGTGGGCGCAGGACAGACGT	1576
Db	42643	TGGCGCGCGACCGGAGAGGCGCTGTGCTGCGCTGTGACCGGCTGGCGCGCGCGGACGCGG	42702
QY	1577	CGCCCGGTGCGGTTCGCAGTATCGCCGATTCCTCAAGCGGCAAGCTCGGCTTTCCTTCA	1636
Db	42703	CGGCGGGCTGTATCCGGGGCACGGTATGCTACGATGCGC--GACCGGGTTCCTTTTCG	42759
QY	1637	CGGACAGAGGGCGCACAGCTGTGGCAATGGGCGGTGGGCTGTACATGTATGATCCGCT	1696
Db	42760	CGGGGACAGGGGCGGACAGCGGTGGGATTTGGGCGGACATGTATCGAGGGGTTCCCGGCT	42819
QY	1697	TTCGCGAAGCGTTTGACCTGTGTGAGGCTGTTCACACAGAGCTTCGACCGGCGCTCC	1756
Db	42810	TGCGCGACCGCCTTGACCACTCCCGCGCGGCTGGACCGGCACTTCGATCTGCTGCTGC	42879
QY	1757	GCGAGGTATGTGGGCGGAAACCGGCGAGCGGTTCGACGCGGCGGCTCGACACAGACGCT	1816
Db	42880	TGCGGGTGTCTTTTTCGCGAGCGGGGATTCGAGCAAGCGCGGCTGTCTGACAGGACCCGGT	42939
QY	1817	TCACCGACCGCGCGCTTTCACCTTCGAAATATGCGCTCCCGCGGCTGTGGCGGATCGTGG	1876
Db	42940	ACGCGACAGCGCGCCTCTTTCGCGGTTCGAGGTGGGTTTTCGACCTGTATCAAGGCTGG	42999
QY	1877	GTTGTAAGCCGGAATTGATGCGCGGCAATAGCATCGTGTAGCTGTGTGGCTGCCTGCTGG	1936
Db	43000	GGGTTCGCGCCGAGTGTCTGTGCGGCACCTCGGTGGCGAGCTGTGGGCGCGGACGTTGG	43059
QY	1937	CGGCGGTGTTTCGCTTGAAGACCGCGGTTTTCGCGGTGCTGTGGCGCGGCGCGCTGATTC	1996
Db	43060	CCGGCTACTGTACACTGAGCAACGCTGTGCGACGTTGTTCGCGGCGCGGCGCGGCTGTATGG	43119
QY	1997	AGGCGCTGCCGCGCGCGGGCGATGTGTGTGATCGAGGCGCGGACGACATGTGTGCTG	2056

[illegible]


```

Db      44178  GCCCTGACCGACCTGCGGGGGGCGCTCGCGCGGCTGTGGTGGGTGACCCGGGGGGGGCGGTGGCC 44237
Oy      3127  TTGTGTAACCACTGACCACTGCGGTGCGCTGCGTGTGTCTGCGCTTCAAGCGCTCTCG 3186
Db      44238  GTGACCGCGCGCGCGAGTCCCGGTGCGTCCGAGACCCAGATTGTGGGGCTTTCGGCGCGTGG 44297
Oy      3187  GCGATGCAAGCGCGCGCGCATGCGCGGTGCGCTGACCTGACCGGACATGTAACCGCTCGACANA 3246
Db      44298  GCGCGCGCTCGAATCTCCCCCGGTCTCTGTGGGGCGGCGCTGTGAGCACTTGCCGTCGACCCGAC 44357
Oy      3247  ATAGCCCGCGCTTCAGCGCGGGGGAGCGGGTGTGATCATGACCGGACCGGCGGGGTGCGT 3306
Db      44358  GACCGGGGACTGTGGGACCTGTGTGGCGGCGCGCTGTGGGGCGCCCGGAGGACCAAGGTGCGGTTC 44417
Oy      3307  CTGCGCGCGGTGTGAGTGGGGCAGACAGCTGGGAGCCGAGGTCCATGCGACGCGCGGACG 3366
Db      44418  CGGGGGGCGGTGCGGTGCGCGCGCGCGGTGTGCTGTCCCG-CGCGCGCGCGCGCGGGCGGAGCG 44476
Oy      3367  CCGGAGAAAGCGCGCTACTGTGAGTGTGCGGTGGGGTGGCGCGGTATGTGAGCGGATTCGCGTGG 3426
Db      44477  GGGCTGTGCGTCCGGCGGCGGACGGTCTGTGTGACCGGGGTGACGAGGCGCGCTCGGCGCGCA 44536
Oy      3427  GACCGGTTTCGTCCGCGCAGTGC 3448
Db      44537  CACGGCCCCGCTGGCTGGCGCGC 44558

RESULT 10
US-10-212-962-2
; Sequence 2, Application US/10212962
; Publication No. US20030114450A1
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Santl, Daniel
; APPLICANT: Myles, David
; APPLICANT: Zq, Tian
; APPLICANT: Hutchinson, Richard
; APPLICANT: Johnson, Robert
; APPLICANT: Zhou, Yi-Qing
; APPLICANT: Feng, Li
; TITLE OF INVENTION: BENZOQUINONE ANZAMYCINS
; FILE REFERENCE: 30062-20075.00
; CURRENT APPLICATION NUMBER: US/10/212.962
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/310.779
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/389.255
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: US 60/393.929
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: No. US20030114450A1 yet assigned
; PRIOR FILING DATE: 2002-07-12
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 6360
TYPE: DNA
; ORGANISM: Streptomyces geldanus
US-10-212-962-2

Query Match      18.0%; Score 765.8; DB 14; Length 6360;
Best Local Similarity 56.1%; Pred. No. 3.2e-159;
Matches 1706; Conservative 0; Mismatches 1237; Indels 98; Gaps 10;

Oy      24  CGCAGCGGAGATCCGATTGCGATCGTGGAGGAGGATGGCCGTGCTCCGGTGGCGGTGAT 83
Db      93  CGAETCCACCGAACCGGTGGGTGATCGTGGAGATGGCGGTGCGCTCCCGGTGGGTGAC 152
Oy      84  CGATCTGAGCGCGGTTCTGAGACGCTCTCGAGGGCTGCGCGACACCGTGGGCGAGTCCC 143
Db      153  CGACCCGGAAGTCGTGTGGAGGCTGTGAGACGAGGGGCGGAGACGATGGGCGCTTCCC 212
Oy      144  CGCCGAAC--GCTGGAGTCAAGACGCTGTGTTGATCCCGACCCCGATGCCCGCGGAGA 200

```

Db	213	CACGGAATCGCGGCTGGGAACCTGGAGACCCTGTTGACTTCGGAATCCGAGCCCGTGGGCA	272
Qy	201	GAGCCCGATTACGCGCGCATCTTTCTTGAGCGAGTGAAGCTGCTTCGACGCGCTTCTT	260
Db	273	GTCTCAAGTACGCGAGGCGGGGTTCTTGAGAGGGGCGGGCGGAATTCAGACGCGCTTCTT	332
Qy	261	CGGCATCTCGCTCTCGCGAAGGCTGCGGATGGAACCTTCACATGACTCTTGTGTGAGGT	320
Db	333	CGGATCTTCGCGCGCGAGAGCCCTGTGTGTGAGACCCGACGAGGGGTGTGTGTGTGAGAC	392
Qy	321	GTGCTGGAGGCGCTGAGAAACGCGCGGATGACTTCATTCGCGCTCTCTGTGTTACGAAAC	380
Db	393	CGCTGGAGAACCTTTCAGAGCGGCGGGGATGGAATCCGCGGTGCTGAGAGGCGGGACAT	452
Qy	381	GGAGGTCTTACGAGATCGGCGCCGTCGACATATGAGAGCGCGCGTCCGACAGCGACGCG	440
Db	453	CGCGGTGTTCGCGGGGGGACGCGGCGACAGGGGTACGCGCGCGGTCCGGGTGA-----GCG	506
Qy	441	GTCCGAGAGATCGACGCTATGCGCGGGCTGAGGAGCAATGCCAGCGCTCGAGCGGGCG	500
Db	507	GCCCAAGGCGCTGAGAGGATTATCTGAGGAGTTCGCGCTTCGCGCAGTGTCAATCTCCGGCG	566
Qy	501	AATCTCGATTGCGCTCGGGGCTGCGAGGGCGGTGTGTGCGCGGTGGAATCGGCTTATGTC	560
Db	567	GTGTCTGTAACGCTTCGAGCTGACCGGATCCGCGGTACGTTGAGACACGCTCTCTGTC	626
Qy	561	CTCGCTGTGCGCTTTATCTGAGCTGTACAGAGCTTGCGCTCGGGGAATGCTTCACGCG	620
Db	627	CTCGCTGTGCGCGCCCATCTCGCGTGAAGAGCGCTGCGGGAATGTTCCATGAGC	686
Qy	621	CTTGGCTGTGGGATTATCGCTGATGTTGTGTGCGGACGACCTCTGTGTGCTTTCGAAGAC	680
Db	687	GCTGGCGGCTGTGTCTCCCGTGAAGGGCGACGCCACCGCTTCTCTCAAGTCTTCCCGGCA	746
Qy	681	CGGGGCGCTGCGCGAGGACGCTGCTGCTCAGAGGCAATTTGCGCGGAGGCGCATGGGTTCCG	740
Db	747	GCGTGGCTTGCGCGCCGACGCGGCGCTGCAAGTCTTTCGCGGGGCGCGCGACGACAC	806
Qy	741	ACGAGGCGAAGGATGCGCGCTGTGTGTCTTCAAGCGGCTGATGAGACCGCGCGGACGCG	800
Db	807	CTGTCTCGAAGGTGTGGGGCTCGTCTGTGTGAGCGGCTGTGCGACCGCGCGCACACG	866
Qy	801	CGATCGAATATTGGCGGTGATTCGAGGATCCGCGATCAATCAAGACGCTGCGACGCGG	860
Db	867	CCAGGAAGTCTGCGCGCTGTGATCCGCGGCGACCGCGGTCAACAGGACGCGCGCTTCAACG	926
Qy	861	TCGTACGCTGCGGAACGGGAGCTCCCAAGAAATGTGTCTGAAACGGGCGCTGCGGAGCG	920
Db	927	ACTCAACGCGCCCAACGCGCCCTTCCAGAGACGGGTATCGCGTAGGCGCTGTTCACAGC	986
Qy	921	AGGCTGCGCGCGCTTCTCGGTGGTTATGTGAGGCAACGCGCACGCGGACGACGCTTGG	980
Db	987	CGGGCTACGCTGGCGGACGCTGAGACGCGCTGTGAGGCCACGCGACCGCGCGCTTCGG	1046
Qy	981	TGACCCCATCGAAATCCAAAGTCTTGAAATGGGATATACGGCTCGGCGGAGATGTGCGCAC	1040
Db	1047	CGACCCCATCGAAACCCAGGCGCGTTCCTGTGCGCACCTTACGCGCCCAAGCCGCGCGAGGCG	1106
Qy	1041	GCGGCTGCTAATCGGGTTCGTTGAGAGCAACCTTGGCGATCTGTAGATATCGTGGGGAT	1100
Db	1107	GCGGCTGTGTGCGGCTTCCCTCAAGTCAACATGCGGACGCGGAGCGCGAGGGGCAAT	1166
Qy	1101	CACGTGGCTGTGAAGGCTGCTTGTCCCTTACAGCACGGGCGAGATTCTGCGGCACTTCA	1160
Db	1167	GCGCAGTGTATCAAGAACGTTCAATGCGCTTACGCGACGGCGGTTGTGCGAAGACCTTCA	1226
Qy	1161	CGCGCAGGCGCTGAACCCCGGATCTGATGGAGGTATCTTGTGCTGACCGTCAAGCGCGC	1220
Db	1227	CGCGGAACAGGCGCACTCCGAGGTGAAGTGAAGCTGCGGCGCGGTGTCTCCGTCTGCGGGA	1286
Qy	1221	CGCGACACCTGTGCGCGGACTGGAATACGCGCGTACGCGCGGGGTGAGCTGTTCGGCAT	1280

Db	1287	GGCGGCGGGCGTGGCCGAGACCGGACACGCCCGCGCGCGCGGGAGATCTCTCTTCGGCGT	1346
OY	1281	GAGCGGGGACCAACCGCGCACGTGGTGTCTGAAAGAAGCGCCGCGCGGAC-----	1328
Db	1347	CAGCGGGGACGAACGACACGTTCATCTTGAAACAGGCCCTCTTAGGGAAGCGAGCGACCGG	1406
OY	1329	-----GTGACACACCGCGCGCGCGAGC	1351
Db	1407	GGAGAACACCGCCGATCAGGAACCGCCCGTACGCTCGCGGGAGTCCGCGAACCCCGGCC	1466
OY	1352	GACCGGCA-----GAGCTGCTGATGTGTCCGCAAGAGACCGCGTACGCTTGGA	1400
Db	1467	GGTGGCCACCGGGCCACGTGGTGGCGGTGGCTGTCTCGGGCCATACGACGAGAGGGCGTGG	1528
OY	1401	TGCACAGCGGGGCGCGCTGCGCGACCAATCTGGAAGCTTACCTTTGCACTGTCTGGCGGA	1460
Db	1527	TGCCAGGCGCGCGCGCTGTGACCCCAAGTGGCGAGACCGCCCTTCGACAGTCCCGCGGA	1586
OY	1461	TGTGGCGTTCAATCTGCGGAGACGACCGCGAGCGGATGAGAGACCGGCGTCCGCGTGGCGGC	1520
Db	1587	CGTGGGCTGTTCATCTAGGCGCACCCCGGACCCGCGTGGACCAACCGCGCGTGTACTTGG	1648
OY	1521	GACGTCGAGGAGAGGGGCTGCGGGGACGCGCTGGACCTCGCGCGGACGAGACAGACGTCCGC	1580
Db	1647	CGCGCATCCCGAGCAGAGGCGCGTCCGCGGGCTGGAGGCGGATGGCTTCGGGCAAGTCCGCCG	1706
OY	1581	CGGTGGGTGGCGCAAGTATCGCCGATTCCTCAACGCGGCAAGCTTCGCTTTCCTTACCGG	1640
Db	1707	GTCGGCGGTTCACCGGGGTCCGTGGGCTC-----CGAAGAAGTGGCGGTTGTTACCGG	1760
OY	1641	ACAGGGGGGCGCAGACGCTGGGCAATGGGCGCTGGGCTGTACATGTATAGTCCGCTTCCG	1700
Db	1761	GCAGGGGACGACGAGCGGGCGGAAATGGGCGCGGACCTGACCGCGCGCCACCGGATTTGGC	1820
OY	1701	CGAGGCGTTGCA-----CCTGTCGTAGGCGCTGTTCAACACGAGAGCTGACCG	1748
Db	1821	GCGGGCTTTCACCGCGCGCTGTGGCGCCAGTTGGGCGACCTGGCGACCGGGAGACGACAGGT	1888
OY	1749	GCCGCTCCGAGAGTGTATGGGCGCGAACCGCGACGCGTGCACCGCGCGCTGTGCACCA	1808
Db	1881	CTCGCTGCGCAGAGTGATCTTCGCGGAGAGGGGTGGCGACGCGACGCGTGTGGACCG	1948
OY	1809	GACAGCTTTCACCCAGCGCGGCGCTGTTCACCTTGCAATATGGCTTCGCGCGCGCTGTGGCG	1866
Db	1941	GACCGAGTTCAACCCAGCGCGCGCTGTTCGCGCTGAGAGTGGCGCTTTCGCGCTGTGGA	2000
OY	1869	GTCGCGGGGTGTAGAGCGCGAGTTGTGTCCGCGGCATATGCAATCCGTGAGCTGTGGCTGG	1928
Db	2001	GTCGTGGGAGATGGCCCGCGCGGTATATGTGTGGGCCACTTCGATCCGCGAGATGGCCGCGC	2060
OY	1929	CTGCGCTGCGGCGCGTGTCTCGGCTTGAAGACCGCGGTGTTCTCGGTGTGGCTGCGCGCGGCG	1988
Db	2061	CCATATGGCCCGGGGTCTGTGCTTCGCCGAGCGCTGCACATTTGGTGGCGGCGCGCGGCGG	2120
OY	1969	CCTGATGACAGGCGCTGCGCGCGCGCGGCGGAGATGTGTTCGATCGAGGCGCGCGAGCGGA	2048
Db	2121	GCTGATGACGACACTCACCGCGCACCGGGGAGCATGTGTCCGCGGTGAGGCGCGCGCGAGAGCA	2180
OY	2049	TGTGGCTGTGCGGTGGGCGGCGCACGACGCTCGGTTCGATTCGCGCGGAGTCAAGCTCC	2108
Db	2181	GATGGCGCGCTGCTCGCGGGGAAAGAGCACAGGTCTTCATTCGCGCGGTATACGGGCC	2240
OY	2109	GGACGAGTGTGCATCGCGGCGCGCGGCAACCCGTGATGCGATCGCGCGCGCGATGGC	2166
Db	2241	GGCGTCCGTGGTGTCTTCGCGGTGACGAGAGACGTGTCAAGCGGTGGGAGAGACGCTGGC	2300
OY	2169	CGCGCGCGGGCGCGACCAAGCGGCTCAAGTCTTCGATGCGTTCACCTACCGCTCAT	2228
Db	2301	GCGGAGGCGCGCAAGACCAAGCGGCTCGTGTCTTCGACGCGCTTCACCTCCCGCCACAT	2360
OY	2229	GGCGCCGATGCTGAGAGGGGCTTCGGGCGGTGTGGCGGAGTCCGAGACTACCGCGCGCGCTC	2286
Db	2361	GGACCGGATGCTGAGCGGCTTTCGAGAGGTGGCGTCCGCGGCTTCGACCTACCGCCACCCCG	2420

OY	2289	GATGTCCTGGATCAGCAATCTGA---GCGGGAAGCTTGGACACAGCAGGTGACTGGC	2345
Db	2421	GATACCCGTGTATCGAACTCACCCGGGCGCGATGCCGATCCCGAGGACTGTCTCCC	2480
OY	2346	GCGCTATTGGGTGGCCCAACGCGCGAGAGTGTCCGCTTCGCGGATGGAGTGAAGCGCT	2405
Db	2481	CGACTACTGGGTATCGGATGACCTGGCGCGCTGCCGTCTCTCCACGGTGTCCACACT	2540
OY	2406	GCAGCGCGCCGAGTCCGGGCACTTCGTCTGAGGTGAGTCCGAAATCGACGCTCTCGGCT	2465
Db	2541	GGCGGACGAGGGCGTGGCCACCTCATCTGGAACTGGGCCCGGATGGGGTGTGACCGCAT	2600
OY	2466	GGTGCCTGGCTGACATGCCGACCGCCCGGCGCTCTCGCATGTTCGCGGCTGGCG	2525
Db	2601	GGGGCAGGACTGTCTGGCCCCGAGCGGACGCGCGCTTGTGTGTCGTCTCCCTGCGCGGCGT	2666
OY	2506	TGAGAGCGCGGCGCACCGTGGCTCGAGGGGCTTCGGGGGCTCTGGGCGGTGGGCTGGT	2585
Db	2661	CCAGAGCGGACGCGGGCTGGGCGGGGCTTCGCGGCTGTACATCGGGGTGTGGGT	2722
OY	2586	CTCTGGGCGCGGCTCTCTCCCC--TCAGGGGGGCGGCGGGTGGCGCTGCCACGTAACC	2642
Db	2721	GGACTGGGAGCGCATGTTCCGCGGGTCCGGGCGCCGGCGCGTCCGCTTCCACGTACGC	2780
OY	2643	TTGGCAGCGCGAGCGCTACTGGATCGACAGAAAGCCGACGACGCGCGCTGGCGACCG	2702
Db	2781	CTTCACAGACGAGCATCTATGGCTGGAGCGGGCGCGGCTCCG-GCGACGTGGGCGCGG	2839
OY	2703	CGGTGCTCCGGGAGCGGGTCAAGACGAGTGAAGAGGGGGCGCGTCCGCGCGGCA	2762
Db	2840	TGGGGCTTCGGGAGAGCGGGCCATCCGCTGCTGGGCGCGGTGTGATGACTCCCGAGAGCG	2899
OY	2763	CGGCGCAGCGCTTCGCTCAACATCCG---CGCCCGAGAGCGGACCGCGGAGAAAG	2818
Db	2900	GCGGGGTGCAAGCTCAGCGGGGCGGTGTCTGTACCGGCGCCCTGTGCTGGGCAACCG	2958
OY	2819	TGAGAGGCGCGCGGAGACCGTCCGTTCCGGGCTCGAGATCGAATGAGCAGAGCGTGTGATC	2878
Db	2960	TCAATCTCGGGGCGGTGTGTGTGCTGCCGACCGCAATGTGTGAATGTGGCTTCGCGCGCG	3019
OY	2879	ACTCTGTCTTCGGGTACGAGAGCGCGCGCCCTGTGCTGGGCGAGGTGAGATTCGCCG	2938
Db	3020	GGGACGAGAACCGGACCCCGGTGCTGGAGGAGCTGTGATCGGGCAGCGCATGTGTGCTGC	3079
OY	2939	TCGACGCGCGCGGGGCTCAGCTTCAATATGTCACGCTCCG	2979
Db	3080	CCGGCGACCCGCTTCAAGTTCAGAGTCTGTGTGGGCGG	3120

RESULT 11
 US-10-156-761-2882
 : Sequence 2882, Application US/10156761
 : Publication No. US20030119018A1
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: OKURA, SATOSHI
 : APPLICANT: IKEDA, HARUO
 : APPLICANT: ISHIKAWA, JUN
 : APPLICANT: HORIKAWA, HIROSHI
 : APPLICANT: SHIBA, TADAYOSHI
 : APPLICANT: SAKAKI, YOSHIYUKI
 : APPLICANT: HATTORI, MASAHIRA
 :
 : TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 :
 : FILE REFERENCE: 249-262
 :
 : CURRENT APPLICATION NUMBER: US/10/156,761
 :
 : CURRENT FILING DATE: 2002-05-29
 :
 : PRIOR APPLICATION NUMBER: JP 2001-204089
 :
 : PRIOR FILING DATE: 2001-05-30
 :
 : PRIOR APPLICATION NUMBER: JP 2001-272697
 :
 : PRIOR FILING DATE: 2001-08-02
 :
 : NUMBER OF SEQ ID NOS: 15109
 :
 : SEQ ID NO 2882
 :
 : LENGTH: 10839


```

Qy 2060 CGGTGGCGCCGCGACGCGTGGGTGATGCGCGGTCAACGCTTCGGACGAGGTGG 2119
Db 6678 -GCTGCGGATGCTTCCGCGGGGTGTGTGCTGCTGCGGTGAACGGGCGCTTCGCTGG 6736
Qy 2120 TCATCGGGGGGCGCGGACACCGCTGCTGATGCGCGCGCGGTATGGCCGCGCGGG 2179
Db 6737 TGCTGTCCGGTGAAGAGGCGGTGACCGCGCTCGCCCAAGCTTGTCTAAACAGGGCA 6796
Qy 2180 CGCGAACCAAGGCGCTCCAGCTTCGCAATGCGTTCACATCAACGCTGATGCGCGGATG 2239
Db 6797 GCGGACCAAGCAAGTGGCGGTCACTGCTTCCTTCATCCGCGGATGAGCCGATGC 6856
Qy 2240 TGAAGCGTTGCGCGGTGCGCGAGTGCATGAGCTACCGCGCGCGGTGATGCTCTGG 2299
Db 6857 TGGCGAGTTGCGGAGACGCTCGCGCGGTGAGATTCCGTCCGCGCGGTATCCCGGTGG 6916
Qy 2300 TCACGATCTGAGGGGAGGCTTGACACAGACGAGTGAAGTCCCGGGCTATTGGGTGC 2359
Db 6917 TGTGGAATGTGACGGGTGGATGGCGGACGCGGATTCACCACTCCGCTTACTGGGTGC 6976
Qy 2360 GCGACGCGCGAGGTGGTGGCTTCGCGATGAGTGAAGGCGCTGACGCGCGCGGTG 2419
Db 6977 GCGATGTGCGGAGGCGCTTGTTCGCGATGATGAGCGACGAGTGTGCGCGGGGTG 7036
Qy 2420 CGGACACCTTGTGATGAGTTCGATCGAATGACGCT-----GCTCGGCTGGTGCCTG 2473
Db 7037 TGAACAGGTTCTGAGCTGGGTTCGCGGGGTGGGCTGACCGCATGGCGAGAGACC 7096
Qy 2474 CTTCGATGCGGAGCGCGCGCGGCTGCTCGCATCTCGCGGCTGAGGCTGAACGAGC 2533
Db 7097 TCGACACACCGCGCGCGACCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7156
Qy 2534 CGGCGACCGTGTCTGAGGCGCTCGGCGGGCTCTGCGGCGGTGAGTGGCTGTCTCTGG 2593
Db 7157 GCGATGCTGACCTTCGCGCTCGCGCTCGGATCAACGCGAGCGGTGCCCGGTTCAGTGT 7216
Qy 2594 CGGCGCTCTTCC---CTCAAGGGGCGCGCGGCTGCGGCTGCGCATGATCCCTTGGAGC 2650
Db 7217 CCGGCGCTGTTGCGCGGACCGGCGCGCGCATCGTTCGATCTCCATCCTCTTCCAGC 7276
Qy 2651 GCGAGCGCTACTGATGACACG---AAAGCCGACGACGCGCGGTGCGACCGCGGTG 2707
Db 7277 ACAAGCGCTACTGCTGCTGCGCACGACCGGACGCGGAGCGGACCGCGCATTCGAC 7336
Qy 2708 CTCGCGGAGCGGTCACGACGAGTTCAGAGAGGGGGCGGTGCGCGCGCGACCG 2765
Db 7337 TGACGCGGACCGGACCGCGCTGCTGCGCGCGGAGTGGCTGCTCCGATCTCGACG 7394

```

RESULT 12

```

US-10-156-761-15102
; Sequence 15102, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OHTURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, NASHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15102
; LENGTH: 125746

```

```

; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15102

```

```

Query Match 17.8%; Score 761.2; DB 14; Length 125746;
Best Local Similarity 57.6%; Pred. No. 2,56-158;
Matches 1589; Conservative 0; Mismatches 1103; Indels 66; Gaps 10;

```

```

Qy 26 CAGCGAAGATCCGATTGCGATCGTCGAGCGAAGTTCGCTCCCGGTGCGGTATG 85
Db 89070 CCGACAGACACCGATGTATGATCGTGGCATGCTGCTCCGCTCCGCGGGTGGCG 89129
Qy 86 ATCTGACGGGTTTGGACCTCTCTGAGGCTTCGCGACACCGTGGGAGTCCCCG 145
Db 89130 GTCCCGAGAGACTTGTGCGTGTCTCGTCAACAGCCCGACGAGATACGGAATCTCCCG 89189
Qy 146 CCGAACGCTGGATGACAGACGCTGTTTATCCCGACCCGATGCCCGCGGAGAACCG 205
Db 89190 CCGACCGAGGCTGGACGCGCTGGCATGAAACGCTTTCATGAGAGATCCGCGAGCC 89249
Qy 206 CCGTTACGCGGCACTTCTCTGAGCGAGTACGCTCTTTCGACGCTCTTCTTCGCA 265
Db 89250 G-----GACAGGCTCTTCTCTGCGCGAGCGCGGCTTCGACGCGGCTTCTTCGGA 89303
Qy 266 TCTGCGCTGCGAAGCGCTGCGAGTGAACCTTCGACATCGACTTGTGAGGTGTGCT 325
Db 89304 TCTGCGCGGAGGACATGAGCATGAGATGATCGCAGACGCGGTGTGCTGAGACCTGCT 89363
Qy 326 GGAAGGCTGAGAGAACCGCGCATGCTCCATGCGGCTCTGTGATGAGAAACGGAG 385
Db 89364 GGAAGGCTGAGAACCGCGCGGTTGATCGGTGCGCTGCGGAGACCGCACCGCG 89423
Qy 386 TGTTCATCGGAGATCGCGCGGCTCCGATATGAGCGCGCTGCGCGCAAGCGCGCTGCG 445
Db 89424 TCTTCGTGCGGCGACACCGCGGAGATGACAGACCTGCTG-----ATGATCTGCTCG 89477
Qy 446 CAGAGATGACGCTCATGAGCGGCTGAGGAGATGCCAGCGTGAAGCGGCGGAACTT 505
Db 89478 AGGACAGCGCGGTTTACGCCCTCACGGGGGTGAGGACGATGATGTCCGCGCGGCTG 89537
Qy 506 CGTATGCTTCGCGGCTGCGAGGCGCTGTGTGCGGTGATACGGCTATTGCTCTGCG 565
Db 89538 CTTATGCTTCGCGGCTGAGAGGCGCGGCTGACAGGTGACACCGCTGCTCTCTGCG 89597
Qy 566 TGTGCGCGGTTTCACTGCGGCTGTCAGAGCTTGGCTCGCGGGAATGCTCCACGCGCTG 625
Db 89598 TGTGACGCTGACCTGACAGGAGCGGCTGCGGCGGCGGATGACCTTGGCGCTG 89657
Qy 626 CTGTTGAGGATATCGCTGATGTTGTGCGCGAGACCTCTGTGTGCTCTGAAAGACCGCG 685
Db 89658 TCGCGGTTGTGACCGCTATGCGGACCGCGGAGGCTTGTGTGATGTTCTGCGGACGGCG 89717
Qy 686 CGTGGCCAGGAGCGGTGCTGCAAGGCAATTTTTCGCGAGGCGGATGGTTCCGACGAG 745
Db 89718 GTCTCGGAGAGACGCGCGGTCAGAGGCTTTCGCGCGCGGTCGAGCGGCGCTGG 89777
Qy 746 GCGAAGGCTGCGCGGTGCTCAAGGCGCTCAATGAGCCCGCGGAGCGGCGATC 805
Db 89778 GCGAGGTTGTGGATCTGCGCTGCGAGGCTTTCGAGACGCGGTGCGGACGGCGCTC 89837
Qy 806 GATATTGCGGATGATTCGAGGATTCGCGATCAATCAACGAGTGCAGACGCGTCTGA 865
Db 89838 CGGTGCTGCGGATGATGCGGATTCGCGGATTCGAGACGAGTGCAGACGCTCTGA 89897
Qy 866 CCGTCCGGAACGGAGCTCCCAAGAAATGCTGCTGAACGCGGCTTGGCGGACGAGCT 925
Db 89898 CCGGCGGAAACGGTCCCTCCAGAGCGGTTGATCCGAGCGGCTGCTGCTGCGCGGCT 89957
Qy 926 GCGCGGCTTGTGCTGATGATGAGGACACCGGACGCGGACGCGGCTTGGTGGACC 985
Db 89958 TGTGCGGCGGAGTGAATGATGAGGCGGACGCGGACGCGGACGCTGCTGCGGACCC 90017
Qy 986 CATTGGAATTCAGGCTGGAATGCGGTATACGGCTTCGCGGCGAGATGTGCGCACGCGCG 1045

```

D	b	90018	CGATCGAGGCAAGGCGTTCTGTGGCAACGTAACGTTCAGAGACCGTCCGGCGGACCCGTC	90077
Q	y	1046	TGCTGATCGGGTCGGTGAAGACCAACCTTGGCCATCCTGAGTATGCGTCGGGGATCATCTG	1105
D	b	90078	TGTGGCTGGGTTGGGTGAGATCGAACATTCGGCCACACCAGTACCGCCGGGTGTGGCG	90137
Q	y	1106	GGCTGCTGAAGAGTGTCTTGTCCCTTACAGACGGGCGAGATTCCTGGGCACTTCCACGGC	1165
D	b	90138	GTTCATCAACAGTCCGTGACTGAGCCCTGTGGCAACGAGCCCTGTTGCCAAGACCTTCGACGTG	90197
Q	y	1166	AGGCGCTGAACCCCGGATCTCAATGGGGGTATCTTCGGCTGAACGCTTCACGGCGCCGGA	1225
D	b	90198	ACAGAGCCGACCCCGAGGTGGACTGGTCCCGGGCCGGCGGGAGCTGTGACGAGAGGGG	90255
Q	y	1226	CACCTTGAGCCGGACTGGAATACGCGCGGACGGGCGGGGGTGAAGCTGTTCCGATGAGCG	1285
D	b	90258	GGGAGTGGCCGGAACCGACGGGGCCGGCCGGCGGGCGGTGTCTTGTTCCGAGTACGG	90317
Q	y	1286	GGAGCAACCGCGCACATGTGTGTGTGAAGAAGCCCGCGCGCGACGTGCAACCGCCG----	1341
D	b	90318	GCAAGAACGGGCACGTATTTCTGAGGACGGCCGCTGTGTCAGGAACGGAACCCCGGTGT	90377
Q	y	1342	--GGCCGGAAGCCACCGGACAGAGCTGTGGTGTGCTGTGCGCAAGAACCGCGTCAAGCCCTG	1399
D	b	90378	CTGCGGTTGACGGGTTCTGTGTCCGTGGGTGGTGTGAGCGCGGTGGAAGAGCCCTTC	90437
Q	y	1400	ATGCAAGAGCGCGCGGCTGCGGACCATTCGAGACCTTACCAAGTGTCTTGGAGCG	1459
D	b	90438	GGGCGACAGGCCCGACCGGCTTGGCGAACGATACGATACGGGGAGACGAT---TGCGGCGGGCG	90499
Q	y	1460	ATGTGGCGTTCAGTTCTGCGCAACGACGCGACGCGATGAGACACCGGCTTCGGGTGGCG	1519
D	b	90495	ACGTGGGTCTGTCTTGTCCGTGTACGCTCGGGTCTGAGACACCGGCGTGTCTTACCG	90554
Q	y	1520	CGACGTGAGGAGGGGCTCGGGCAACCCCTGACGCTGCGGAGGAGGAGACAGACGTTCG	1579
D	b	90555	GCGGGGATCGGAGAGACTTCTGTGTCAGATTGGCCGGGTTGGAGAGGGTGGCGGGCGG	90614
Q	y	1580	CCGGTGGCGTCCGAGTATCGCGGATTCCTCAACGCGGCAAGCTCGGCTTCTTTCACCG	1639
D	b	90615	CGGGGATGTGTGCGGGGTGTGGCGGGAG-----GGCGGGAGGGGTTCTGTTACCG	90668
Q	y	1640	GACAGGGGGCGGACGCTGTGGCATGGGCGGTGGGCTGTACGATGATGTGCTCCGCTTCC	1699
D	b	90669	GTCAAGGCTGTGCGACGCGGCTTGTATGGGGGTGAGAGCTGTACGACGCTTCCCGTGTTCG	90728
Q	y	1700	GCGAGGGGTTGCACCTGTGGGTAGAGGCTGTCAACCAAGAGGCTGCACCGGCGGCTCGCG	1759
D	b	90729	CGGCTGGGTTGCAGGGGGGTGTGTGCGAGTTGAGTCTGTTGTGAGAGCGGCGGTTAAG	90788
Q	y	1760	AGGTGATGTGGGCGGAACCGGACGAGCTGCACGCGCGGCGTCTGACCAAGACAGCTTCA	1819
D	b	90789	AGGTGCTTCAACCGA-----TGCGCGGGTGTGACCGGACCGTTCATCA	90833
Q	y	1820	CCAGCGCGGCGTGTTCACCTTGATATACGCTCGCGCGCTGTGGCGGCTGTGGGGTG	1879
D	b	90834	CGCAGGCGGGTGTGTTGCGCTTGAGAGGTGCGGCTGTTCAGAGCTGTGTGGGTTGCTGTGGGGTG	90893
Q	y	1880	TAGAGCGCGAGTGTGTGCGCGGCGCATAGCATCGGTGACGTGTGTGTGCTGCTGTGTGGCGG	1939
D	b	90894	TGCGGGGAGATGTGCTCTGAGGAGATTGCATCGGTGAGCTGGGCGGGCGGCTGATGTGGCGG	90953
Q	y	1940	GCGTGTCTGCGTTGAAGACGCGGTTGTTCTGTGTGGCTGTGCGCGCGCGGCGCTTATCAG	1999
D	b	90954	GGGTGTGTGCTGTGCGGATGTGTGTGCGCGGATGTGTGGGCGGGCGGGCGGTTGATGACAG	91013
Q	y	2000	CGCTGCGCGCGCGGGGCGAGTGTGTGCATCGAGCGCGGAGCGCGATGTGGCTGTG	2059
D	b	91014	CGCTGCGGAGAGGGGCTGTGCATGTGCTGTGCGAGGGCGGCGAGGACGA-----	91064
Q	y	2060	CGGTGGGCGGCGACGACGCTGTGTGCATCGCCGGGTTCACACGCTCCGGAACCAAGTGG	2119

Db	91063	GGCTGCGCGGTCTCCCGGGGGGTGTGTCGATGGCTGCGGTGAACGGGGCCCGTTGCTGTG	91122
Oy	2120	TCAATCGCGGCGCCCGGGGCAACCCGTGTGATTCGATCGCGCGGTGATGTCGCCGCGGGGG	2179
Db	91122	TGCTGTCTCGGTGTATGAGAGACCCTGTATACCGCGCTGTCCCAAGCGTTTGTGAACAGGGCA	91181
Oy	2180	CGGAACCAAGAGCGCTTCACAGTGTTCGATATGCTTCATCTACCGCTCATGAGCCCGATGC	2239
Db	91182	GGCGGACCAAGCAAGCTTGGCGGTATAGTACAGCCCTTTCACACTCCGCGCGGATGGAGCCGATCC	91241
Oy	2240	TGAGAGCGTTCCGAGCGCTGTGGCCGAGTCCGATGAGCTAACCGCGGCGCGTGTGATGTCTTGG	2299
Db	91242	TGGCGGAGTTCCGCGGAGACGCTTCGCCGCGGTGAGTTCCGTCCGCCGCTATCCCGGTGG	91301
Oy	2300	TGACGATCTGACCGGGGAAAGCGCTTGACAGACGAGGTGAGCTGCCGGGCTATTGGGTGTC	2359
Db	91302	TGTGAAATGTAGCGGGGTGGGATGTGGGAGACCGCGAGTTCACTCCGCGCTACTGGGTTC	91361
Oy	2360	GCCACCGCGAGAGAGTGTGTGCGCTTTCGCGGATGTGAGTGAAGGCGCTGCACGCGGCGGTG	2419
Db	91362	GGCATGTGCGGAGAGGCGCTTTCGTTTGCGCGATGTGTGTGGAGAGGTGCTGGCGGGGGTG	91422
Oy	2420	CGGGCACTTTCGTGTGAGGTGTGTTCCGAATCCAGCCT-----GCTCGGCTGTGTTCCTG	2473
Db	91422	TGGACAGAGTTCTTCGAGCTGTGGTCCGCGGGGTGCGCTGACCGCGAGATGGCCGAGAACCC	91481
Oy	2474	CTTGCATGTCCGAGACGCGCCCGGCGCGGCTGTGCATGTGTGCGCGCTGTGGGCTGTGACGAC	2533
Db	91482	TTCACACACACCGGCGCGCGAGCGCTGTGTGCTGCTGCTGCTGACATCCCAACCCCGGAG	91541
Oy	2534	CGGCGACCGTGTCTGAGAGCGCTTCGCGGGGCTGTGGCGCTCGGTGGCGCTGTCTCTGGG	2593
Db	91542	CGGACTGCTGACCCCTTCGCGCTGTGGCGGATCCACCGAGCGGTGCCCCGCTGTGACTGTG	91601
Oy	2594	CCGGCCCTCTTCCC---CTACGGGGGGGGGGGGGTGTGCGTGTGCCAATGATCCCTTGGGACG	2650
Db	91602	CCGGGCGTGTGCGCGGACCGGCGCCCGGACCGTCCGAACTGTCCCACTTACGCTTTCACG	91661
Oy	2651	GCGAGCGCTTACTGATTCGACAGC--AAAGCCGACGACCGGAGCGGTGCGACCGCGGTG	2707
Db	91662	ACAAAGCGCTTACTGGCTGGACTGTGCGGACGACCGGCGAGCGGACCCCGGAGCATCGGAC	91721
Oy	2708	CTCCGGAGCGGGTCAACGAGAGGTGTGAGAGAGGGGGCCCGGTGCGCGGCGCGACCG	2765
Db	91722	TGACCGGAGACCGGACACACCGGTGCTGTGGCGCGGAGTGTGCGCTGCCGACTCGGAGG	91779

```

RESULT 13 '761-1/c
US-10-156-761-1/c
Sequence 1, Application US/10156761
Publication No. US20030119018A1.
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, KAZUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, NASHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:

```


Db 3560536 -GCTCCGCTGCTTCCCGGGGATGTCTGCTGCTGCTGAACGGGCCCCGCTTGGCTGG 3560538
Qy 2120 TCATCGCGGGCGCGGGGCAACCGTGCATGCGATGCGGGCGGATGCGCGCGCGG 2179
Db 3560537 TGCTCTCCGGTATGAGGAGCGCGTGAACCGCTCGCCCAAGCGTTTGTGAACAGGGA 3560478
Qy 2180 CGCGAACCAAGGCGCTCCACGCTTCCGATGCGATGCGTTCACCTCACTCAATGCGCCGATTC 2239
Db 3560477 GCGGAGCAAGAGTTGGCGGTGATGACGCTTCCACTCCGCGGATGAGAGCGGATTC 3560418
Qy 2240 TGGAGCGCTTGGCGGCTGTGGCGGATGCGATGCGTGAAGTCAACCGCGCGCTGCTGCTG 2299
Db 3560417 TGGCGGAGTTCCCGGAGAGCTTCGCGCGGATGAGTTCGCTCCGCGCTATCCGGTGG 3560358
Qy 2300 TCAGCAATCTGAGCGGGAAGGCTTGACACAGAGAGTGAAGTCCGCGGCTATTTGGTTC 2359
Db 3560357 TGTCCGAATGTGACGGGTGGGATGGGCGAAGCGGAGTTTCACTCCCGCTACTGGGTTC 3560298
Qy 2360 GCCAGCGCGGAGAGGTGTGGCTTCCGCGGATGAGTGAAGGCGGTGACAGCGCGCGGCTG 2419
Db 3560297 GGCATGTGGGAGGCGCTTCTTTCGCGGATGTGTGGCGAAGGCTGGCGCGGCTGG 3560238
Qy 2420 CGGGACCTTCTGAGAGTGGTTCGAAATCGACGCT-----GCTCGGCTGTGCTG 2473
Db 3560237 TGGACAGGTTCTCTGAGCTGGGTTCCGCGGGTGGCGCTGACCGGATGGCGGAGAGACC 3560178
Qy 2474 CTGTGATGCGGAGCGCGCGCGGCTGCTGCAATGTCGCGCGCTGGGCGTGAAGAC 2533
Db 3560177 TCGACCAACACCGGCGCGCGCGCTGTGCTGCTGCTGCTGCAACCGCGGAG 3560118
Qy 2534 CGGCAACGCTGCTGAGAGGCTTCGCGGCTTGGGCGCTGCTGCTGCTGCTGCTG 2593
Db 3560117 CGGACTGCTGACCTCTCGGCTTCGCGGATCGACGAGCGGTGCGCGGCTGCTGCT 3560058
Qy 2594 CGGCTCTTCTCC---CTCAGGAGGAGCGGCGGCTGCGCTGCGCAAGTACCTTGGAC 2650
Db 3560057 CGGGCTGTTCGCGGACCGCGCGCGCGCGCGCGCTGCGCACTGCGCTTCCAGC 3559998
Qy 2651 GCGAGCGCTACTGATTCGACAG---AAGCGGACGAGCGGCGGCTGCGACCGCGCTG 2707
Db 3559997 ACAACGCTACTGCTGCTGCTGCGCGACCGGCGAGCGGCGGACCGCGGACATCGAC 3559938
Qy 2708 CTCGCGGAGCGGCTGACGAGGCTGAGAGGCGGCGGCTGCGCGCGCGACCG 2765
Db 3559937 TGAAGCGGACCGGACCGCGCTGCTGCGCGCGGAGTGGCGGCTGCCCACTCGACGG 3559880

RESULT 14
US-09-861-289-1
Sequence 1, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/861,289
CURRENT FILING DATE: 2001-05-18
PRIORITY APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FaSTSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 15872
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 17.4%; Score 742; DB 10; Length 15872;
Best Local Similarity 55.7%; Pred. No. 5e-154;

Matches 1594; Conservative 0; Mismatches 1215; Indels 53; Gaps 7;
Qy 26 CAGCCGAAGATCCGATTCGATTCGTGAGAGGAGTTCGCTGCTGCGGCTGGCTGATTC 85
Db 8490 CCGACGAGATCCGATTCGATTCGTGAGAGGAGTTCGCTGCTGCGGCTGGCTGATTC 8549
Qy 86 ATTCGAGGAGGTTTCGAGGCTCCCTCGAGGCTCGCGGACACGTCGAGGAGTCCCG 145
Db 8550 CGCCGAGAGACTGTGAGGCTGTGAGGCTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8609
Qy 146 CGAAC---GCTGGAGTACAGAGGCTGTTGATCCGACCCGATGCCCCGGGAGGA 202
Db 8610 TCAACCGGCGGAGGAGCTGAGAGGCTTCAAGACCGGATCCGAGTCCGAGTCCGAGCA 8669
Qy 203 CGCCGTTAGCGCGCATCTTCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 262
Db 8670 GGTACTGCGGAGGAGGCGGAGGCTTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 8729
Qy 263 GCATCTGCTGCGGAGGAGGCTGCGGATGAGACCTGACATCGACTTCTTGTGAGGTGT 322
Db 8730 GCATCTGCTGCGGAGGAGGCTGCGGATGAGACCTGACATCGACTTCTTGTGAGGTGT 8789
Qy 323 GCTGGAGGAGGCTGAGAGAGCGCGGATGCTGCTGATCGGCGCTGCTGCTGCTGCTG 382
Db 8790 CTTGGAGGAGGCTGAGAGAGCGCGGATGCTGCTGATCGACCGGCTGCTGCTGCTGCTG 8849
Qy 383 GAGTGTATGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 442
Db 8850 GTGTCTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8909
Qy 443 CGGAGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 502
Db 8910 GCTGAGAG---GCTATCTGCTGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8963
Qy 503 TCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 562
Db 8964 TCTCTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9023
Qy 563 CGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 622
Db 9024 CGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9083
Qy 623 TGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 682
Db 9084 TGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9143
Qy 683 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 742
Db 9144 AGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9203
Qy 743 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 802
Db 9204 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9263
Qy 803 ATTCGATATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 862
Db 9264 ACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9323
Qy 863 TGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 922
Db 9324 TGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9383
Qy 923 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 982
Db 9384 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9443
Qy 983 ACCCATGGAATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1042
Db 9444 ACCCATGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9503
Qy 1043 CGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1102
Db 9504 CGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9563

Qy 1103 CTGGCTGCTGAAGTCTCTTGTCTTCCTTACGACGAGGAGATTCTTGCACCTTCACG 1162
Db 9564 GCGGCTCATCAAGATGCTCCAGAGGCTGCGGCAAGCTTGTGCTCCGACCTTCACG 9623
Qy 1163 GCGAGGCTTAACCCCGGATCTCATAGGGTGAATCTTGGCTGACCGTCAAGCGCGCC 1222
Db 9624 CCGAGAACCCAGCCCGATGCTGAGCTCCGCGCGGATACGGCTGTCTCATCTCCG 9683
Qy 1223 GGAACACCTGAGCTGGAATATACGCGGAGGCGGGGATGAGCTGTTCCGATGA 1282
Db 9684 AGGTGCTTGGCAGAGGACCGGCGGCTCCGCGGACCGGAGTGTCCCTTCGCGTCCG 9743
Qy 1283 GCGGAGCAACGCGGACGCTGCTGGAAGAGCGCGCGCGGACGTCGACACCGCGCG 1342
Db 9744 GCGGACCAATGCCATGTCTGCTCGAAGAGGACCGCGCGCGCGCGGACCGG 9803
Qy 1343 GCGGAGGACGAC-----GGCAGAGCTGCTGG 1369
Db 9804 CCGGAGGAGCGCGCGGCTCCGCGCGGACGAGAGGCGGAGGCGCTGCGCTGGG 9863
Qy 1370 TGCTGTGGCAAGACCGGCTGACGCTTGAATGACAGAGGCGCGGCTGGCGACATC 1429
Db 9864 TGTCTCCGAGCGGACGAGCGGCGCTGCGGCTCCAGGACCGCGGCTCCGCGACCC 9923
Qy 1430 TGAGACCTTACCTTTCGAGTGTCTGAGGAGTGTGGCTTCAATCTGCGGACGACGCA 1489
Db 9924 TCTCCGAGACCGCGGCGCGCGCGCGCGGATGCTCTTCTCCCTCCGCGCGACGCGG 9983
Qy 1490 GCGGATGAGAGACCGGCTCGGAGTGGCGGCGGACGTCGAGGAGGAGGCTGCGGCGCGC 1549
Db 9984 CAGGCTTTGACACCGCGCGCTGATCGGCTCGGAGCGGCGCGACCTCCGCGCGCC 10043
Qy 1550 TGAGCGCTGCGGCGGAGGAGACAGACGTGCGCGGCTGCGGCTGCGATTCGCGATTCCT 1609
Db 10044 TGAGCGCTTGGCGGAGGAGCGGACGAGTCCGCGGCTGCGGAGTTCGCGCA---CC 10100
Qy 1610 CAGCGGCAAGCTGCTCTTCTTTCACCGGACAGGAGGCGGACGCTGCGGATGAGCGC 1669
Db 10101 GGGAGCGGAGGATGCGCTTCTCTTTCACCGGAGGAGGAGGACGCGCGCGGAGTGGCC 10160
Qy 1670 GTGGCTGTACGATGATGATGCTCGGCTTCGCGGAGGCTTGAATCTGCGTGAAGCTGT 1729
Db 10161 ACGACCTTACGCTGCGGCTTCTTTCGCGTTCGCGCTGCGCTGCGAGTGAACGCGCTC 10220
Qy 1730 TCAACGAGAGCTGACCGGCGGCTTCGCGGAGTGAATGAGGCGGACCGGCGAGCTGCG 1789
Db 10221 TCGACCGGCTGCTGCGGCGGCGGCTGCGGCGGCTGCGAGCGCGGACCGGCTGCGCG 10280
Qy 1790 ACGCGGCTGCTGACGAGACGCTTACCCAGGCGGCGCTGTTCACTTTCGATATG 1849
Db 10281 AAGCGGCACTCTGAGACCGGACCGAGTACACCGGCGGCTTTCGCGCTTCAGAGTGG 10340
Qy 1850 CGCTGCGGCGCTGCGGCTGCGGCTGAGAGCGGAGTGTGCTGCGCGGCGCATGACA 1909
Db 10341 CGCTCAACGCGCTGCTGAGGACCTGCGGAGTGCCTGCTGCGGCGCATGCG 10400
Qy 1910 TCGGAGCTGTGCTGCTGCTGCGGCGGCTGCTGCTGAGAGCGGCTGCTGCTGCTGCTG 1969
Db 10401 TGGGGAATGAGCGGCGGCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10460
Qy 1970 TGTGCTGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2029
Db 10461 TGTGCTGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 20520
Qy 2030 TCGAGGCGGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2089
Db 10521 TCGCGGCGGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10580
Qy 2090 TCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2149
Db 10581 TCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10640

Qy 2150 CGATCGCGGCGGAGTGGCGGCGGCGGCGGAGCCGAGCCCTTCACCTTCGATG 2209
Db 10641 AGCGGCGGCGGAGCTGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10700
Qy 2210 GCTTCCACTACCGCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2269
Db 10701 CTTTCACTACCGCTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10760
Qy 2270 TGAAGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2326
Db 10761 TGGCTTACCGGAGACCGGAGCTGAGGCTGCTTCAAGTTCACGCTTACGCGGCGGCGG 10820
Qy 2327 CAGAGAGTGAAGTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2386
Db 10821 CCGGTGAATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10880
Qy 2387 CCGATGAGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2446
Db 10881 CCGAGGCGGCTCGGAGGAGCAGCAGCTTGGAGGCGGCGGCGGCGGCGGCGGCGG 10940
Qy 2447 AATGAGCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2506
Db 10941 ACGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11000
Qy 2507 CATGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2566
Db 11001 ---CGGCGATCAACAGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11057
Qy 2567 GGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2626
Db 11058 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11115
Qy 2627 CGTGCACGCTTACCTTGGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2686
Db 11116 CGAGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11175
Qy 2687 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2746
Db 11176 CCGCTTACCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11235
Qy 2747 CCGTGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2806
Db 11236 CCGGAGCGGCTGCGGCTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11295
Qy 2807 GCGGAGAGGCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2848
Db 11296 CCACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11337

RESULT 15
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860.846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1
Query Match 17.4%; Score 742; DB 10; Length 15872;

Best Local Similarity 55.7%; Pred. No. 5e-154; Matches 1594; Conservative 0; Mismatches 1215; Indels 53; Gaps 7;

```
OY 26 CAGCCGAGATCGATTGCGATGTCGAGCAGAGTTCCCTGCTCCCGTGCATGCG 85
Db 8490 CCGAGACCATCGATCGGATGTCGATGCGCATGCGCATGCGGATGCGGATGCGG 8549
OY 86 ATCTAGAGGCGGTTCTGAGCGCTCTCGAGAGGCTCGGCGCAGACCCGTCGGGCGAGTCCCG 145
Db 8550 CCGCGAGGACCTGTGCGCGCTGTGTGCGAGAGGCGAGACCGGATTCACGAGTTCCCG 8609
OY 146 CCGAAC---GCTGGATGAGCAGCAGCTGTTGATCCGACCCCGATGCGCGGAGAGA 202
Db 8610 TCACCGCGGCTGGGACCTGAGAGCCTTACGACCCGATCCGAGTGAAGGAGACA 8669
OY 203 CGCCGCTTACGCGCGCATTTCTCTGAGCGAGCGTAGCCCTTTCGACGCTCTTCTTTCG 262
Db 8670 CCGTACTGCGCGGAGGCGGCTTCTGGAAGCGCGGATGATTCGACGCGCTTCTTTCG 8729
OY 263 GATCTCGGCTCGGAGAGGCTCGGATGAGACCTTCGACATGACCTTTCGAGAGTGT 322
Db 8730 GATCTCGGCTCGGAGAGGCTCGGATGAGACCTTCGACATGACCTTTCGAGAGTGT 8789
OY 323 GCTGGAGGCGCTGAGAGACCGCGCATGCTTCATCGCGCTCGCTCGGATGAGAGAG 382
Db 8790 CCGGAGGCGCTGAGAGACCGCGCATGCTTCATCGCGCTCGCTCGGATGAGAGAG 8849
OY 383 GAGTGTTCATCGGAGATCGCGCGCTCGGATGAGAGCGCGCTCGGAGAGAGCGCT 442
Db 8850 GTGTCTACCTGGGCGCGCGCATGCTTCATCGCGCTCGGATGAGAGAGCGCT 8909
OY 443 CCGAGAGATGACCGCTCATGCGCGGCTGGGAGACATGCGCAGCGTGGAGGCGGAG 502
Db 8910 GCTCGAG---GAGTATGCTGACCGGAGCGCGGAGCGCGGATGCTCGGCGGAG 8963
OY 503 TCTCTATGCTCGCGGCTCGAGAGCGCTGTGTGCTCGGATGAGATGCGCTATGCTCT 562
Db 8964 TCTCTATGCTCGCGGCTCGAGAGCGCTGTGTGCTCGGATGAGATGCGCTATGCTCT 9023
OY 563 CGCTGTGCGCTTTCATGCTGCTGCTGAGAGCTTTCGCGGATGAGAGAGCTTTCAG 622
Db 9024 CGCTGTGCGCTTTCATGCTGCTGCTGAGAGCTTTCGCGGATGAGAGAGCTTTCAG 9083
OY 623 TGGGCTGAGGCTGATGCGGATGCTGTGCGGAGACCTTCGATGAGCTTTCGAGAGCC 682
Db 9084 TGGGCTGAGGCTGATGCGGATGCTGTGCGGAGACCTTCGATGAGCTTTCGAGAGCC 9143
OY 683 GGGCGCTGCGCAGGAGCGCTGCTGAGAGCAATTTTCGCGGAGGCGGATGAGAGCT 742
Db 9144 AGGGCTGCGCGCGCAGGAGCGCTGCTGAGAGCAATTTTCGCGGAGGCGGATGAGAGCT 9203
OY 743 GAGCGGAGAGGCTGCGCGCTGCTGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 802
Db 9204 GGGCGGAGAGGCTGCGCGCTGCTGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 9263
OY 803 ATCGAGATATGCGGCTGATTCGAGAGATCGCGGATGAGATGAGAGCTTTCGAGAGCT 862
Db 9264 ATCGAGATATGCGGCTGATTCGAGAGATCGCGGATGAGATGAGAGCTTTCGAGAGCT 9323
OY 863 TGAACGTCGCGAGAGAGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 922
Db 9324 TGAACGTCGCGAGAGAGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 9383
OY 923 GCTGCGCGGCTTTCGAGAGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 982
Db 9384 GCTGCGCGGCTTTCGAGAGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 9443
OY 983 ACCCGATGAGATTCAGAGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 1042
Db 9444 ACCCGATGAGATTCAGAGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 9503
OY 1043 CGCTGCTGAGAGGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 1102
Db 1102 CGCTGCTGAGAGGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 1102
```

```
Db 9504 CGCTGCTGAGAGGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 9563
OY 1103 CTGGGCTGCTGAGAGGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 1162
Db 9564 CGGCGGATGAGATTCAGAGGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 9623
OY 1163 CGGAGGCTGAGAGGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 1222
Db 9624 CGGAGGCTGAGAGGCTTTCGAGAGCAATTTTCGCGGAGGCGGATGAGAGCTTTCAG 9683
OY 1223 GAGACCGCTGCGGATGAGATTCAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1282
Db 9684 AGGTCCTGCGGATGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 9743
OY 1283 GGGGAGCGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1342
Db 9744 GGGGAGCGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 9803
OY 1343 CGCGGAGCGAGC-----GAGAGAGCTGCTG 1369
Db 9804 CGCGGAGCGAGC-----GAGAGAGCTGCTG 9863
OY 1370 TCTGTGCGAGAGAGCGGCTGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1429
Db 9864 TGTCTGCGAGAGCGGCTGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 9923
OY 1430 TGGAGACTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1489
Db 9924 TCTCCGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9983
OY 1490 GCGGATGAGAGAGCGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1549
Db 9984 CAGCTTTCGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10043
OY 1550 TGAAGCTGCGGCGGAGAGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1609
Db 10044 TGAAGCTGCGGCGGAGAGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10110
OY 1610 CAGCGGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1669
Db 10101 GAGAGGAGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10160
OY 1670 GTGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1729
Db 10161 GTGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10220
OY 1720 TCAACGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1789
Db 10221 TCAACGAGAGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10280
OY 1790 AGCGGCGCTGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1849
Db 10281 AGCGGCGCTGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10340
OY 1850 CGCTGCGGCTGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1909
Db 10341 CGCTGCGGCTGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10400
OY 1910 TGGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 1969
Db 10401 TGGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10460
OY 1970 TGGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 2029
Db 10461 TGGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10520
OY 2030 TGAAGCGCGGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 2089
Db 10521 TGAAGCGCGGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10580
OY 2090 TCGCGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 2149
Db 10581 TCGCGGCTGAGAGGCTTTCGAGAGGCTTTCGAGAGGCTTTCAG 10640
```

Oy	2150	CGAATCGGCGGCGCATGATGCGCGCGCGGCGGAAACAAGGCGCTCACGTTGCGCATG	2209
Db	10641	AGCGCGCGCGCGCACTGCCGCGAGAGGCCGCGCACACGAGCGATTCGCGTCCGCGACG	10700
Oy	2210	CGTTCCATCAACCGCTCATGAGCCCGCATGCTGAGAGCGTTTCGGGCGTGTGGCCGAGTCGG	2269
Db	10701	CGTTCCACTCAACCCCTGATGAGACGGCATGTCTCCGGAATTCGGGAGGTGCGCGCCGGCC	10760
Oy	2270	TGAGCTACCGCGCGCGCGTGCATGCTCTGTGTCAGCAATCGACCGGAGG---GCTTGA	2326
Db	10761	TGCGCTACCGGGAACCGAGACTGACGCTGCTTTCACCGGTACAGGGGCGGCGCCCGCC	10820
Oy	2327	CAGAAGAGTGAAGCTCCGCGCGCTATTGGGTGCGCCACGCGCGAGAGTGTGCGCTTCG	2386
Db	10821	CGGCTGAATCAACCGGCGCCGCACTAGGATGCGCCAGGTCGATGACCCGTGCGCTTCG	10880
Oy	2387	CGGATGGAATGAAAGCGCTGCAACGCGCGCGGTCCGGGCACCTTCTGTGAGATGCTGCGCA	2446
Db	10881	CGGACGCGGTCCGCGACACGACCGGCTCTGGAACCCGCGACCTTCTGTGAGACCTGGCCCG	10940
Oy	2447	AATGACGCTGCTTGCGCGCTGTGCTCCTGACATGCGCGAGCGCCCGCGCGCGCTGTCG	2506
Db	10941	ACGCGGTGCTGTGTGGGATGAGAGAGATGCTCTGAGAGACGACACCGTGGCTGTGCTGC	11000
Oy	2507	CATTGTGCGCGCGCTGGGCGTGAAGAGCCGCGCAACCTGTCTCGAGCGCTTCGCGCGCTCT	2566
Db	11001	---CGGCGATTCACAAACCCCGGACACCGCGCTCGACGATCGGCGGCTCTCCCGCGCGCTGC	11057
Oy	2567	GGGCGGTGCGGTGCGCTGTCTCTGCGGGCGCGGCTCTTCCCTCAAGGGGGCGCGCGGTGCG	2626
Db	11058	GGGCGGCGCGCGCGCTGACGCGCGGGGCGCCCGCGGTGAGCT---GGGCGGATGACGCG	11115
Oy	2627	CGCTGCCACGTACCTTTCGACGCGCGAGCGCTACTGATTCGACAGAAAGCCGACGACG	2686
Db	11116	CGAGGCGCCGAGGAGGCGCGGCGCGCGCTGCAATGCGCGGTCCAGCGCTTCGCGGACCG	11175
Oy	2687	CGGCGCGGTGCGAACCGCGGTGCTTCGCGGAGCGGCTCAACAAGAGTTCAGAGGGGGCG	2746
Db	11176	CGGCTTACTGGTTCGCCCCCGGGGCGCGCGCGGACACGAGACTGGATTTACCGGATTCGG	11235
Oy	2747	CGGTGCGCGCGCGGACCGGCGGACGAGCGCTTGGGCTTCGACATTCGCCCGCCGAGAGCGAC	2806
Db	11236	CTGGGACCGGCTGCGGCTGTGACCGGCGGAGGCCGGAACCGCGCGGCTGCTGTGTGAT	11295
Oy	2807	GCCGGAGAAAGTTCGAGAGCCCGCGGCGCAACCTGTCCGTTCCGGC	2848
Db	11296	CCACCCGACAGCCCGCGCTGCGCGGAGCTGTCTCCGCGACGCG	11337

```
Search completed: October 4, 2003, 12:39:42
Job time : 1051.7 secs
```

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 3, 2003, 11:46:26 ; Search time 225.999 Seconds.
(without alignments)
8331.640 Million cell updates/sec

Title: US-10-014-717-1_COPY_7610_11875

Perfect score: 4266
Sequence: 1 GTGGCGATGCTCCATCGA.....GGGAATCATGCGCATGA 4266

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4266	100.0	68750	3 US-09-335-409-1	Sequence 1, Appl1
2	4266	100.0	68750	4 US-09-568-102-1	Sequence 1, Appl1
3	4266	100.0	68750	4 US-09-567-969-1	Sequence 1, Appl1
4	4266	100.0	68750	4 US-09-568-480-1	Sequence 1, Appl1
5	4266	100.0	68750	4 US-09-568-486-1	Sequence 1, Appl1
6	4266	100.0	68750	4 US-09-568-472-1	Sequence 1, Appl1
7	4266	100.0	68750	4 US-09-567-899-1	Sequence 1, Appl1
8	4187.6	98.2	71989	4 US-09-443-501A-2	Sequence 2, Appl1
9	814.2	19.1	3978	4 US-09-266-965-19	Sequence 19, Appl1
10	814.2	19.1	12249	4 US-09-266-965-19	Sequence 74, Appl1
11	814.2	19.1	18331	4 US-09-266-965-96	Sequence 96, Appl1
12	803.2	18.8	33529	4 US-09-144-085-3	Sequence 3, Appl1
13	755.2	17.7	43280	2 US-08-804-227C-1	Sequence 1, Appl1
14	742	17.4	15872	3 US-09-105-537-1	Sequence 1, Appl1
15	740.8	17.4	4411529	3 US-09-103-840A-1	Sequence 1, Appl1
16	732	17.2	4403765	3 US-09-103-840A-2	Sequence 2, Appl1
17	725	17.0	49377	1 US-08-764-233A-1	Sequence 1, Appl1
18	712.4	16.7	32155	4 US-08-311-731A-1	Sequence 1, Appl1
19	701.2	16.4	80161	3 US-09-036-987A-1	Sequence 1, Appl1
20	701.2	16.4	80161	3 US-09-370-700-1	Sequence 1, Appl1
21	701.2	16.4	80161	3 US-09-603-207-1	Sequence 1, Appl1
22	675.2	15.8	44377	2 US-08-804-227C-7	Sequence 7, Appl1
23	675.2	15.8	44377	2 US-08-804-198-1	Sequence 1, Appl1
24	671.8	15.7	28958	1 US-08-258-261B-6	Sequence 6, Appl1
25	671.8	15.7	28958	1 US-08-456-837-6	Sequence 6, Appl1
26	671.8	15.7	28958	1 US-08-457-342-6	Sequence 6, Appl1
27	671.8	15.7	28958	1 US-08-457-646A-6	Sequence 6, Appl1

28	671.8	15.7	28958	1 US-08-458-076A-6	Sequence 6, Appl1
29	671.8	15.7	28958	1 US-08-764-233A-4	Sequence 4, Appl1
30	671.8	15.7	28958	1 US-08-457-335A-6	Sequence 6, Appl1
31	671.8	15.7	28958	1 US-08-729-214-6	Sequence 6, Appl1
32	671.8	15.7	28958	3 US-09-028-934-6	Sequence 6, Appl1
33	663.2	15.5	20235	1 US-07-642-734C-3	Sequence 3, Appl1
34	663.2	15.5	20235	3 US-08-439-009A-3	Sequence 3, Appl1
35	662.2	15.5	13987	2 US-08-804-227C-13	Sequence 13, Appl1
36	649.4	15.2	47981	4 US-09-410-551B-1	Sequence 1, Appl1
37	633	14.8	77536	4 US-09-410-551B-1	Sequence 1, Appl1
38	630.8	14.8	11219	3 US-07-642-734C-1	Sequence 1, Appl1
39	630.8	14.8	11219	3 US-08-439-009A-2	Sequence 2, Appl1
40	628.8	14.7	4403765	3 US-09-103-840A-2	Sequence 1, Appl1
41	628.8	14.7	4411529	3 US-09-103-840A-1	Sequence 30, Appl1
42	627.6	14.7	13842	3 US-09-105-537-5	Sequence 5, Appl1
43	627.6	14.7	36778	3 US-09-105-537-5	Sequence 19, Appl1
44	627.6	14.7	38506	3 US-09-320-878-19	Sequence 1, Appl1
45	627.6	14.7	38506	4 US-09-141-908-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-335-409-1
Sequence 1, Application US/09335409
Patent No. 6121029
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OR INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match	100.0%;	Score 4266;	DB 3;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4266;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GTGGCGATGCTCCATCGAGCGCGGAGCGCGGAGATCCGATTCGATCGGAGCGAGT	60	
DB	7610	GTGGCGATGCTCCATCGAGCGCGGAGCGCGGAGATCCGATTCGATCGGAGCGAGT	7669	
QY	61	TGCGCTGCGCGCGATGATGATCTGAGCGGTTCTGAGCGCTCTCGAGGCTCG	120	
DB	7670	TGCGCTGCGCGCGATGATGATCTGAGCGGTTCTGAGCGCTCTCGAGGCTCG	7729	
QY	121	CGGACACCGTGGGAGATGCTCCCGCGGAAAGCTGGGATGACAGAGCGTGTGATCC	180	
DB	7730	CGGACACCGTGGGAGATGCTCCCGCGGAAAGCTGGGATGACAGAGCGTGTGATCC	7789	
QY	181	GACCGGATGCGCGGAGGAGAGCGCGGATGAGCGGATCTTCGAGGAGTACC	240	
DB	7790	GACCGGATGCGCGGAGGAGAGCGCGGATGAGCGGATCTTCGAGGAGTACC	7849	
QY	241	TGCTTGACCGCTCTCTTCTGCGATCTGCGGAAAGCGTGGGATGAGACCTGCA	300	
DB	7850	TGCTTGACCGCTCTCTTCTGCGATCTGCGGAAAGCGTGGGATGAGACCTGCA	7909	
QY	301	CATGACCTCTGCTGAGAGTGTCTGAGAGCGGCTGGAAGACCGCGATGCTCATCG	360	
DB	7910	CATGACCTCTGCTGAGAGTGTCTGAGAGCGGCTGGAAGACCGCGATGCTCATCG	7969	

QY	361	TCGCTCGCTCGGTAACGAAACGGAGAGTTCATCCGGATACGGGCGCTCCGAATATAGGC	420
Db	7970	GGCTCTGTCGCTACGGAAACGGAGATTTCATCGGATACGGGCGCTCCGAATATAGGC	80223
QY	421	GGCTGCTCGCAAGCGACGGCGCTCGCAGAGATCGACGCTCATGCGGGCTGGGACGATG	480
Db	8030	GGCTGCTCGCAAGCGACGGCGCTCGCAGAGATCGACGCTCATGCGGGCTGGGACGATG	80893
QY	481	CCGACGCTCGGAGCGGGCGGAAATCTCGATGCGCTCGGGCTGCGAGGGCGGTGTGCGG	540
Db	8090	CCGACGCTCGGAGCGGGCGGAAATCTCGATGCGCTCGGGCTGCGAGGGCGGTGTGCGG	81493
QY	541	GTGGAATACGAGCCTATTGCTCTCGCTCGCTGGGCGCGCTTCATCTGAGCTTGACAGCTTGGC	600
Db	8150	GTGGAATACGAGCCTATTGCTCTCGCTCGCTGGGCGCGCTTCATCTGAGCTTGAGCCTTGGC	82093
QY	601	TCCGGGGAAATGCTCCACGAGCCCTGGCTGTGGGGATATCGCTGATGTTGTCCCGAGCAC	660
Db	8210	TCCGGGGAAATGCTCCACGAGCCCTGGCTGTGGGGATATGCTGATGTTGTCCCGAGCAC	82693
QY	661	CTCGTGTGGCTCTCGAAGACCCGGGCGGTGGCCAGGGACGGGTGCTGAGAGGCAATTTTG	720
Db	8270	CTCGTGTGGCTCTCGAAGACCCGGGCGGTGGCCAGGGACGGGTGCTGAGAGGCAATTTTG	83223
QY	721	GGCGAGCGCGATGAGTGGTTCGAGACGAGGCGAAGGGTGCCTGTGTGTCTCAAGCGGCTC	780
Db	8330	GGCGAGCGCGATGAGTGGTTCGAGACGAGGCGAAGGGTGCCTGTGTGTCTCAAGCGGCTC	83893
QY	781	AGTGAAGCCCCCGCGGACCGCGATTCGGATATTGGCGGTGATTTCAAGATCCGCCATTCAT	840
Db	8390	AGTGAAGCCCCCGCGGACCGCGATTCGGATATTGGCGGTGATTTCAAGATCCGCCATTCAT	84493
QY	841	CACGACGGTGGAGACGAGCGGTCTGACCGTGCAGAACGGAGACTCCCAAGAAATCGTCTG	900
Db	8450	CACGACGGTGGAGACGAGCGGTCTGACCGTGCAGAACGGAGACTCCCAAGAAATCGTCTG	85093
QY	901	AAACGGGCGCTTGGCGGACCGACGAGCTGCGCCCGCTTTCGGTGGTTATGTGAGGACAC	960
Db	8510	AAACGGGCGCTTGGCGGACCGACGAGCTGCGCCCGCTTTCGGTGGTTATGTGAGGACAC	85693
QY	961	GGACACGGGCGACGAGCGCTTGTGACCCCAATCGAAATCCAAGCTCTGAAGCGGTATACGGC	1020
Db	8570	GGACACGGGCGACGAGCGCTTGTGACCCCAATCGAAATCCAAGCTCTGAATGCGGTATACGGC	86223
QY	1021	CTCGGGCGAGATGTGCGCCAGCGCCGCTGCTGATCGGGTGGTGAAGACCAACCTTGGCCAT	1080
Db	8630	CTCGGGCGAGATGTGCGCCAGCGCCGCTGCTGATCGGGTGGTGAAGACCAACCTTGGCCAT	86893
QY	1081	CCTGAATATGCTGTGGGGATCACTTGGGCTGCTGAAGGTGCTTCTCCCTTACGACCGG	1140
Db	8690	CCTGAATATGCTGTGGGGATCACTTGGGCTGCTGAAGGTGCTTCTCCCTTACGACCGG	87493
QY	1141	CAGATTCTCTGCGCACTTCACGCGCAGCGCGCTGAAACCCCGGATCTCATGAGGATGATTT	1200
Db	8750	CAGATTCTCTGCGCACTTCACGCGCAGCGCGCTGAAACCCCGGATCTCATGAGGATGATTT	88093
QY	1201	CGGCTGACCTGTACCGCGCGCCCGGACACCTGTGGCCGACCTGAAATACCGCCGACGCGCG	1260
Db	8810	CGGCTGACCTGTGTACCGCGCGCCCGGACACCTGTGGCCGACCTGAAATACCGCCGACGCGCG	88693
QY	1261	GGGGTGAAGCTGTTCGAGATAGAGGGGACCAAGCGCGACGTGTGCTGGAAGAAGCGCGCG	1320
Db	8870	GGGGTGAAGCTGTTCGAGATAGAGGGGACCAAGCGCGACGTGTGCTGGAAGAAGCGCGCG	89223
QY	1321	GGCGCGACGTGCAACACCGCGCGCGCCGAGCGACCGGCGAGAGCTGTGTGCTGTCCGCA	1380
Db	8930	GGCGCGACGTGCAACACCGCGCGCGCCGAGCGACCGGCGAGAGCTGTGTGCTGTCCGCA	89893
QY	1381	AGGACCCGGGTCAAGCCTCTGGAATGCAACAGCGGCGCGGCTGTGCGGACATCTGAGACCTAC	1440
Db	8990	AGGACCCGGGTCAAGCCTCTGGAATGCAACAGCGGCGCGGCTGTGCGGACATCTGAGACCTAC	90493

QY	1441	CTTTCGAGAGTCTGAGGCGAATGTGGCGTTAGTCTGGCGACAGACGGCAGCGCATGGAG	1500
Db	9050	CTTTTCGAGAGTCTGGGCGAGTGTGGCTTACAGTCTGGCGAGAGCCAGCGCATGGAG	9109
QY	1501	CACCGGCTCGCGGAGTGGCGCGACCTTCAGAGGAGGGGCTGGCGGCGCCTTGAACGTGCG	1560
Db	9110	CACCGGCTCGCGGAGTGGCGCGACCTTCAGAGGAGGGGCTGGCGGCGCCTTGAACGTGCG	9169
QY	1561	GCGCAGGGACAGAGCTGCGCCCGGTGGCGGTGGCGACATTCGCCGATTCTCACCGCGCAAG	1620
Db	9170	GCGCAGGGACAGAGCTGCGCCCGGTGGCGGTGGCGACATTCGCCGATTCTCACCGCGCAAG	9229
QY	1621	CTCCGCTTTCTTTCACCGGACAGGGGGGCGCAGACGCTGGGGCAATGGGCGGTGGGCTGTAC	1680
Db	9230	CTCCGCTTTCTTTCACCGGACAGGGGGGCGCAGACGCTGGGGCAATGGGCGGTGGGCTGTAC	9289
QY	1681	GATGTATGTGTCGCGGTTCCGCGAGGCGTTGCGACTGTGTGCGTGAAGCTGTTCACACAGAG	1740
Db	9280	GATGTATGTGTCGCGGTTCCGCGAGGCGTTGCGACTGTGTGCGTGAAGCTGTTCACACAGAG	9349
QY	1741	CTCGACCGCGCGGCTCGCGAGGTGATGTGGGCGGAAACGGGCGACGCTGACCGCGCGCTG	1800
Db	9350	CTCGACCGCGCGGCTCGCGAGGTGATGTGGGCGGAAACGGGCGACGCTGACCGCGCGCTG	9409
QY	1801	CTCGACCAAGACAGCTTTCACCCAGCGCGCGCTGTTCACCTTCCAAATATGTGCTGTCGCGCG	1860
Db	9410	CTCGACCAAGACAGCTTTCACCCAGCGCGCGCTGTTCACCTTCCAAATATGTGCTGTCGCGCG	9469
QY	1861	CTGTGTGGCGGATGTATGGGGGTATAGAGCCGGAGTTGGTTCGCGGACATACATTCGGTGAAGCTG	1920
Db	9470	CTGTGTGGCGGATGTATGGGGGTATAGAGCCGGAGTTGGTTCGCGGACATACATTCGGTGAAGCTG	9529
QY	1921	GTGGCTGCTGTCGCTGTGGCGGGCGGTTCCTGCTTGAGAGCGCGGTTCTGTGTGGCTGCG	1980
Db	9530	GTGGCTGCTGTCGCTGTGGCGGGCGGTTCCTGCTTGAGAGCGCGGTTCTGTGTGGCTGCG	9589
QY	1981	CGCGGCGCGCTGATATGCAAGGCGCTGCCCGCGCGGGCGCAATGTGTGTGATATCGAGCGCGCG	2040
Db	9590	CGCGGCGCGCTGATATGCAAGGCGCTGCCCGCGCGGGCGCAATGTGTGTGATATCGAGCGCGCG	9649
QY	2041	GAGGCGCGATGTGATGTGCTGCGAGTGGCGCGCACAGCACTGTGCGGTGATGCGCGCGGCT	2100
Db	9650	GAGGCGCGATGTGATGTGCTGCGAGTGGCGCGCACAGCACTGTGCGGTGATGCGCGCGGCT	9709
QY	2101	AACGCTCCGAGCAAGATGTATGCGCGGGCGCGGGCGAACCCGTGCATGTGCATGCGGCGCG	2160
Db	9710	AACGCTCCGAGCAAGATGTATGCGCGGGCGCGGGCGAACCCGTGCATGTGCATGCGGCGCGCG	9769
QY	2161	GCGATGTGCGCGCGCGGGGCGCGAAACCAAGCGCTTCAGCTTCTGCAATGTGCTTCACTCA	2220
Db	9770	GCGATGTGCGCGCGCGGGGCGCGAAACCAAGCGCTTCAGCTTCTGCAATGTGCTTCACTCA	9829
QY	2221	CCGCTCATGTGCCCCGAGATGTGAGGGCGTTGGGGGTGTGGCCGAGTGGAGTACCGG	2280
Db	9830	CCGCTCATGTGCCCCGAGATGTGAGGGCGTTGGGGGTGTGGCCGAGTGGAGTACCGG	9889
QY	2281	CGGCGCTGCAATCGTCTGTGTACAGCAATCTGAGCGGAGAGCTTTCACACAGAGGTGAGC	2340
Db	9890	CGGCGCTGCAATCGTCTGTGTACAGCAATCTGAGCGGAGAGCTTTCACACAGAGGTGAGC	9949
QY	2341	TGCGCGGGCTATATGTGGGTGCGCACGCGCGAGAGTGTGCGCTTCGCGATGTGATGTAAG	2400
Db	9950	TGCGCGGGCTATATGTGGGTGCGCACGCGCGAGAGTGTGCGCTTCGCGATGTGATGTAAG	1000
QY	2401	GCGCTGCACCGCGCGGGTGGCGGACCTTGTGTGAGAGTGGAGTCCGAAATGACAGCGCTGCTC	2460
Db	10010	GCGCTGCACCGCGCGGGTGGCGGACCTTGTGTGAGAGTGGAGTCCGAAATGACAGCGCTGCTC	1006
QY	2461	GCGCTGTGTGCTGCTGCATATGCGGAGCGCCCGGCGCGGCGCTGCATGCTGCGCGGCT	2520
Db	10070	GCGCTGTGTGCTGCTGCATATGCGGAGCGCCCGGCGCGGCGCTGCATGCTGCGCGGCT	1012
QY	2521	GGGGCTGATCGAGCGGCGGACCGTCTGAGAGCGCTCGGCGGGCTGTGGGCGGTGCGGTGCG	2580

10130 GGGGGTACAGCCGCGGACCGTCTGAGCGCTCCGCGGCTCTGAGCGCTGAGCGG 10189
2581 CTGGTCTCTGAGCGCGGCTCTTCCCTCAGAGGGGGGGGGGGGGGGGGGGGGGGGGGG 2640
10190 CTGGTCTCTGAGCGCGGCTCTTCCCTCAGAGGGGGGGGGGGGGGGGGGGGGGGGGGG 10249
2641 CTTTGGAGCGCGGACCTGATCGACGAGAAAGCCGACGAGCGCGCGCTGGGCGAC 2700
10250 CTTTGGAGCGCGGACCTGATCGACGAGAAAGCCGACGAGCGCGCGCTGGGCGAC 10309
2701 GCGCGTCTCTGAGCGCGGCTCTGAGCGCTGAGCGCGCGGCGGCGGCGGCGGCGG 2760
10310 GCGCGTCTCTGAGCGCGGCTCTGAGCGCTGAGCGCGCGGCGGCGGCGGCGGCGG 10369
2761 GACCGGCGGACCGCTCGGCTGACATCCGCGCGCGGAGCGGCGGCGGAGAGGTC 2820
10370 GACCGGCGGACCGCTCGGCTGACATCCGCGCGGAGCGGCGGCGGAGAGGTC 10429
2821 GAGCGCGCGGCGGACCGCTCGGCTGAGATCGATGAGCGGCGGCTGATGATCAC 2880
10430 GAGCGCGCGGCGGACCGCTCGGCTGAGATCGATGAGCGGCGGCTGATGATCAC 10489
2881 CTGGTCTCTGAGCGCGGCTCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2940
10490 CTGGTCTCTGAGCGCGGCTCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10549
2941 GACCGGCGGCGGCGGCTCGGCTGAGATCGATGAGCGGCGGCGGCGGCGGCGGCGG 3000
10550 GACCGGCGGCGGCTCGGCTGAGATCGATGAGCGGCGGCGGCGGCGGCGGCGGCGG 10609
3001 CTGGCGGAGAAAGCCCAACCTCTCGGCTCTGAGCGCGGCGGCGGCGGCGGCGGCGG 3060
10610 CTGGCGGAGAAAGCCCAACCTCTCGGCTCTGAGCGCGGCGGCGGCGGCGGCGGCGG 10669
3061 GCGGTGGGCGAGGGCGTGAAGCGCTCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 3120
10670 GCGGTGGGCGAGGGCGTGAAGCGCTCTGAGCGCGGCGGCGGCGGCGGCGGCGGCGG 10729
3121 GAGCGGCTTCTACCGGCTGACCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3180
10730 GAGCGGCTTCTACCGGCTGACCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10789
3181 CTCTCGGCGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3240
10790 CTCTCGGCGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10849
3241 GACGAAATAGCCCGCTTCAAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3300
10850 GACGAAATAGCCCGCTTCAAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10909
3301 GTGGGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3360
10910 GTGGGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10969
3361 GGGACGCGCGGAGAAAGCGGCTGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3420
10970 GGGACGCGCGGAGAAAGCGGCTGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11029
3421 GCGTCCGAGCGGCTGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3480
11030 GCGTCCGAGCGGCTGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11089
3481 GTGCTCAACCTCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3540
11090 GTGCTCAACCTCGGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11149
3541 GGGCGGCTTGTGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3600
11150 GGGCGGCTTGTGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11209
3601 CCGTTCGCGCAATCTCTCTCTGCTGAGTCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 3660

11210 CCGTTCGCGCAATCTCTCTCTGCTGAGTCTCGGCGGCGGCGGCGGCGGCGGCGG 11269
3661 CCGGCGGCGGCTCGGCGGCTCTGAGAGGCTCTCGGCGGCGGCGGCGGCGGCGGCGG 3720
11270 CCGGCGGCGGCTCGGCGGCTCTGAGAGGCTCTCGGCGGCGGCGGCGGCGGCGGCGG 11329
3721 ACCCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3780
11330 ACCCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11389
3781 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3840
11390 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11449
3841 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3900
11450 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11509
3901 CTGGCGTCAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3960
11510 CTGGCGTCAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11569
3961 GTCTCGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4020
11570 GTCTCGGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11629
4021 CTGGCGTCAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4080
11630 CTGGCGTCAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11689
4081 CTGAGCTGTCAGCGAGCTTCTGTCAGGCTCCCGCAATATCGCTTGTGGCCAAAC 4140
11690 CTGAGCTGTCAGCGAGCTTCTGTCAGGCTCCCGCAATATCGCTTGTGGCCAAAC 11749
4141 CTGTTGATGCTTCTGCGCACAGCTCTCTCTTGAAGCGGCTGCGGCGGCGGCGGCGGCGG 4200
11750 CTGTTGATGCTTCTGCGCACAGCTCTCTCTTGAAGCGGCTGCGGCGGCGGCGGCGGCGG 11809
4201 GCGAGCGTGTCAAAAGCACTTCTGTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4260
11810 GCGAGCGTGTCAAAAGCACTTCTGTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 11869
4261 CTATGA 4266
11870 CTATGA 11875

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goelach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-568-102-1

Query Match 100.0%; Score 4266; DB 4; Length 68750;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGCGGATCGTCCCATGAGCGCGACGCCAAGATCCGATTCGATGTCGAGCGAGT 60
Db 7610 GTGCGGATCGTCCCATGAGCGCGACGCCAAGATCCGATTCGATGTCGAGCGAGT 7669
QY 61 TGCGCTCTGCCCCGCTGGGCTGATCGATTGAGCGGGTTCGAGCGCTCTCGAGGGCTCG 120
Db 7670 TGCGCTCTGCCCCGCTGGGCTGATCGATTGAGCGGGTTCGAGCGCTCTCGAGGGCTCG 7729
QY 121 CGCGACACCGTCTGAGCGAGTCCCCGCCGACGCTGGATGACAGCGTGGTTGATCCC 180
Db 7730 CGCGACACCGTCTGAGCGAGTCCCCGCCGACGCTGGATGACAGCGTGGTTGATCCC 7789
QY 181 GACCCCGATGCCCCGGGAGAGACGCCCGTTACGCGCGATCTTCTGAGCGAGTACCC 240
Db 7790 GACCCCGATGCCCCGGGAGAGACGCCCGTTACGCGCGATCTTCTGAGCGAGTACCC 7849
QY 241 TGCTTCAGCGCTCTTCTTGCGCATCTCGCTCGCGAAGCGCTGCGGATGAGACCTGCA 300
Db 7850 TGCTTCAGCGCTCTTCTTGCGCATCTCGCTCGCGAAGCGCTGCGGATGAGACCTGCA 7909
QY 301 CATGCACTCTTGTGAGAGTGTGCTGGAGGCGCTGAGAGACCGCGCATGCTCAATCG 360
Db 7910 CATGCACTCTTGTGAGAGTGTGCTGGAGGCGCTGAGAGACCGCGCATGCTCAATCG 7969
QY 361 GCGCTCGTGGTACGAAAACGAGAGTGTTCATCGGAGATCGGCCCGTCCGAATATAGGCC 420
Db 7970 GCGCTCGTGGTACGAAAACGAGAGTGTTCATCGGAGATCGGCCCGTCCGAATATAGGCC 8029
QY 421 GCGCTCGCGAAGGACGCGCTCCGCAAGATTCACGCTCATGCGCGCTGGGACCATG 480
Db 8030 GCGCTCGCGAAGGACGCGCTCCGCAAGATTCACGCTCATGCGCGCTGGGACCATG 8089
QY 481 CCCAGCGTCTGAGGCGGCGCAATCTGATGCTCCCTCGGGCTGCGAGGGCCGTGTGTCG 540
Db 8090 CCCAGCGTCTGAGGCGGCGCAATCTGATGCTCCCTCGGGCTGCGAGGGCCGTGTGTCG 8149
QY 541 GTGATATCGGCTATTCGTCTCGTGGTGCCTTCATCTGCGCTGTCAAGCTTGGCG 600
Db 8150 GTGATATCGGCTATTCGTCTCGTGGTGCCTTCATCTGCGCTGTCAAGCTTGGCG 8209
QY 601 TCCGGGGAATGCTCAAGCGCTGCTGGTGGGGGATGCGTGAATCTTCCCGACGACC 660
Db 8210 TCCGGGGAATGCTCAAGCGCTGCTGGTGGGGGATGCGTGAATCTTCCCGACGACC 8269
QY 661 CTGCTGTGCTCTGCAAGACCGGGCGCTGCGCAAGGACGCTGCTGCAAGGCAATTTTCG 720
Db 8270 CTGCTGTGCTCTGCAAGACCGGGCGCTGCGCAAGGACGCTGCTGCAAGGCAATTTTCG 8329
QY 721 GCGAGGCGCATGAGTTTGAAGAGCGAAGGATGCGCGCTGTGTCTTCAAGCGCTC 780
Db 8330 GCGAGGCGCATGAGTTTGAAGAGCGAAGGATGCGCGCTGTGTCTTCAAGCGCTC 8389
QY 781 AGTGAAGCGCGCGGAGACGCGGATTCGATATTTGGGCGGATTCAGAGATCCGCAATCAAT 840
Db 8390 AGTGAAGCGCGCGGAGACGCGGATTCGATATTTGGGCGGATTCAGAGATCCGCAATCAAT 8449
QY 841 CACGACGCTGAGCAGCGCTGTAACCTGCGCAACGAGGACCTCCCAAGAAATCGTCTG 900
Db 8450 CACGACGCTGAGCAGCGCTGTAACCTGCGCAACGAGGACCTCCCAAGAAATCGTCTG 9059
QY 901 AAACGAGGCTCTGGGAGACGAGGCTGCGCGCGTCTTCTGTTGGTTATGTGAGGACAC 960
Db 8510 AAACGAGGCTCTGGGAGACGAGGCTGCGCGCGTCTTCTGTTGGTTATGTGAGGACAC 8569
QY 961 GGCACGCGCAGAGCTTGTGATCCCATCGAAATCCAGCTGGAAGCGGTATACGCG 1020
Db 8570 GGCACGCGCAGAGCTTGTGATCCCATCGAAATCCAGCTGGAAGCGGTATACGCG 8629
QY 1021 CTCGGGCGAGATGTTCGACCGCGCTGTGATCGGGTCTGTAAGACCAACTTGGCAT 1080
```

```
Db 8630 CTCGGGCGAGATGTTCGACCGCGCTGTGATCGGGTCTGGTGAAGACCAACTTGGCAT 8689
QY 1081 CTTGATATGCTGTGGGATATCATGCGCTGTGAAGTCTGTTGTCTCTTACAGCGG 1140
Db 8690 CTTGATATGCTGTGGGATATCATGCGCTGTGAAGTCTGTTGTCTCTTACAGCGG 8749
QY 1141 CAGATTCTCGGACCTTCAAGCGGAGCGCTGAACCCCGGATCTCATGAGGGATCTT 1200
Db 8750 CAGATTCTCGGACCTTCAAGCGGAGCGCTGAACCCCGGATCTCATGAGGGATCTT 8809
QY 1201 CGGCTGACCGTCAAGCGCGCCCGGACACCGTGGCCGACTGGAATACGCGCGA 1260
Db 8810 CGGCTGACCGTCAAGCGCGCCCGGACACCGTGGCCGACTGGAATACGCGCGA 8869
QY 1261 GGGGTGAGCTGTTCGCAATGAGCGGAGCAACCGCGCATGTGTCTTGAAGAGCGCG 1320
Db 8870 GGGGTGAGCTGTTCGCAATGAGCGGAGCAACCGCGCATGTGTCTTGAAGAGCGCG 8929
QY 1321 GCGGCGACGTGACACCGCGCGCGCGGAGCGAACCGGACAGAGCTGTGTGCGCA 1380
Db 8930 GCGGCGACGTGACACCGCGCGCGCGGAGCGAACCGGACAGAGCTGTGTGCGCA 8989
QY 1381 AGGACCGCTGACGCTTGAATGCAACAGCGCGCGCGCTGCGACCATCTGAGACCTTAC 1440
Db 8990 AGGACCGCTGACGCTTGAATGCAACAGCGCGCGCGCTGCGACCATCTGAGACCTTAC 9049
QY 1441 CTTTCCGAGTGTCTGGGCGATGTGCGCTTCAATCTGGCGACAGCGCGACGCGATGAG 1500
Db 9050 CTTTCCGAGTGTCTGGGCGATGTGCGCTTCAATCTGGCGACAGCGCGACGCGATGAG 9109
QY 1501 CACCGGCTCGCGGAGGCGGAGCGTGAAGGAGGAGGCTGCGGAGCGCTGAGACGCTGCG 1560
Db 9110 CACCGGCTCGCGGAGGCGGAGCGTGAAGGAGGAGGCTGCGGAGCGCTGAGACGCTGCG 9169
QY 1561 GCGCAGGAGACAGAGTGCCTCGGTCGCTGCGAGTATCGCCGATCTTCAACGCGGCAAG 1620
Db 9170 GCGCAGGAGACAGAGTGCCTCGGTCGCTGCGAGTATCGCCGATCTTCAACGCGGCAAG 9229
QY 1621 CTGCGCTTTTCTTTCAACCGGACAGGAGGCGCAGACGCTGGGCAATGGGCTGTGTAC 1680
Db 9230 CTGCGCTTTTCTTTCAACCGGACAGGAGGCGCAGACGCTGGGCAATGGGCTGTGTAC 9289
QY 1681 GATGTATGTCCTCGGCTTCCGCGAGGCGTTCGACCTGTGCGTGAAGGCTGTCAACAGAG 1740
Db 9290 GATGTATGTCCTCGGCTTCCGCGAGGCGTTCGACCTGTGCGTGAAGGCTGTCAACAGAG 9349
QY 1741 CTCGACCGGCGCTCCGCGAGAGTATGTGAGCCGGAACCGGCGAGCGCTGAGCGCGCTG 1800
Db 9350 CTCGACCGGCGCTCCGCGAGAGTATGTGAGCCGGAACCGGCGAGCGCTGAGCGCGCTG 9409
QY 1801 CTGACCCAGACAGCTTCAACCCAGACCGCGCTGTTCATCTTCCAAATATGCGCTCGCG 1860
Db 9410 CTGACCCAGACAGCTTCAACCCAGACCGCGCTGTTCATCTTCCAAATATGCGCTCGCG 9469
QY 1861 CTGTGCGGCTGTGTGAGGCTGTAGAGCCGAGATGGTTCGCGCGACCATAGCATCGGTGAGCTG 1920
Db 9470 CTGTGCGGCTGTGTGAGGCTGTAGAGCCGAGATGGTTCGCGCGACCATAGCATCGGTGAGCTG 9529
QY 1921 GTGCGTCCCTGTGCGGCGCGCTGTCTGCTTGAAGACGCGGTTCTGTGTGCTGCG 1980
Db 9530 GTGCGTCCCTGTGCGGCGCGCTGTCTGCTTGAAGACGCGGTTCTGTGTGCTGCG 9589
QY 1981 GCGGAGGCGCTGATGACAGGCGCTGCGCGCGCGCGGCGATGTGTGATTCAGAGCGCG 2040
Db 9590 GCGGAGGCGCTGATGACAGGCGCTGCGCGCGCGCGGCGATGTGTGATTCAGAGCGCG 9649
QY 2041 GAGGCGATGAGCTGTGCGGCTGCGCGCAACGACGTCGAGTGTGATTCGCGCGCTC 2100
Db 9650 GAGGCGATGAGCTGTGCGGCTGCGCGCAACGACGTCGAGTGTGATTCGCGCGCTC 9709
QY 2101 AACGCTTCGAGACAGATGTATTCGCGGAGCGCGGAGCAACCGGTGACATGCGAGCG 2160
Db 9710 AACGCTTCGAGACAGATGTATTCGCGGAGCGCGGAGCAACCGGTGACATGCGAGCG 9769
```

QY	2161	GGGATGAGCCGCGCGGGGGCGGAAACCAAGCGCTCCACGTCTCGATCGCTTCCACTCA	2220
Db	9770	GCGATGCGCGGCGCGGGGCGCGAAACCAAGCGCTCCACGTCTCGATCGCTTCCACTCA	9829
QY	2221	CCGCTCAATGGCCCCGGAATGCTGGAAGCGCTTGGGCGGTGGCCGAGTCGGTGAACCTACCGG	2280
Db	9830	CCGCTCAATGGCCCCGGAATGCTGGAAGCGCTTGGGCGGTGGCCGAGTCGGTGAACCTACCGG	9889
QY	2281	CGGCGCTCGAATCGCTCTGCTCAGCAACTTGAGCGGGAGAGCTTGGCAACAGACGAGTGAGC	2340
Db	9890	CGGCGCTCGAATCGCTCTGCTCAGCAACTTGAGCGGGAGAGCTTGGCAACAGACGAGTGAGC	9949
QY	2341	TCGCGCGGAGCTAATGTTGGTGGTCGCAACGCGCGAGAGGTGGTCGCTTCGCGGATGGAGTGAAG	2400
Db	9950	TCGCGCGGAGCTAATGTTGGTGGTCGCAACGCGCGAGAGGTGGTCGCTTCGCGGATGGAGTGAAG	10009
QY	2401	GCGCTGCAACGCGCGGCTGCGGGGACACTTGTGTGAAGTGGGTCCGAATTCGACGCTGGTC	2460
Db	10010	GCGCTGCAACGCGCGGCTGCGGGGACACTTGTGTGAAGTGGGTCCGAATTCGACGCTGGTC	10069
QY	2461	GGCGCTGGGCGCTGCTGCATAGCGGGAAGCCCGGCGCGGCGCTGCTGCATCGTCGCGCGCT	2520
Db	10070	GGCGCTGGGCGCTGCTGCATAGCGGGAAGCCCGGCGCGGCGCTGCTGCATCGTCGCGCGCT	10129
QY	2521	GGGCGGTGACGAGCCGCGGAGACCGTGTCTGAGAGCGCTCGGCGGAGCTCTGGGCGGTGGGCT	2580
Db	10130	GGGCGGTGACGAGCCGCGGAGACCGTGTCTGAGAGCGCTCGGCGGAGCTCTGGGCGGTGGGCT	10189
QY	2581	CTGGTCTTCCTGAGGCTCGGCTCTTCCCTCAAGGGGGGGGCGGTGGCTGCTGCCACGTAC	2640
Db	10190	CTGGTCTTCCTGAGGCTCGGCTCTTCCCTCAAGGGGGGGGCGGTGGCTGCTGCCACGTAC	10249
QY	2641	CCTTGGGACGCGCGGACGCTACTGGATGAGACAGAAAGCGGACGACGCGCGCGGTGGCGAC	2700
Db	10250	CCTTGGGACGCGCGGACGCTACTGGATGAGACAGAAAGCGGACGACGCGCGCGGTGGCGAC	10309
QY	2701	CGCGGTGCTCCGAGAGCGGGTCAAGAGAGTGAAGAGAGGAGGCGCGGTGCGCGCGAC	2760
Db	10310	CGCGGTGCTCCGAGAGCGGGTCAAGAGAGTGAAGAGAGGAGGCGCGGTGCGCGCGAC	10369
QY	2761	GACCGGCGGACGCTCGGCTCGACCATCGGCTCGGACGCGGACGCGGAGAAAGTCT	2820
Db	10370	GACCGGCGGACGCTCGGCTCGACCATCGGCTCGGACGCGGACGCGGAGAAAGTCT	10429
QY	2821	GAGGCGCGCGCGGACCGCTCCGTTCCGCTCGGAATGAGTGAAGCGACGCTGTATATAC	2880
Db	10430	GAGGCGCGCGCGGACCGCTCCGTTCCGCTCGGAATGAGTGAAGCGACGCTGTATATAC	10489
QY	2881	CTCGGTGCTTGGGTCAGGAGCGGCGGCGCTGCTGGGCGAGGTCAGATCGGCGTC	2940
Db	10490	CTCGGTGCTTGGGTCAGGAGCGGCGGCGCTGCTGGGCGAGGTCAGATCGGCGTC	10549
QY	2941	GACGCGGCGGAGCTCAGCTTCAATGATGTCAGCTCGCGCTGGGACATGGTCCGACGAC	3000
Db	10550	GACGCGGCGGAGCTCAGCTTCAATGATGTCAGCTCGCGCTGGGACATGGTCCGACGAC	10609
QY	3001	CTGCGCGGAAAGCCCAACCTCTCGCTGTGCTGAGAGCGAGTGGCGCGGCGCATCGTC	3060
Db	10610	CTGCGCGGAAAGCCCAACCTCTCGCTGTGCTGAGAGCGAGTGGCGCGGCGCATCGTC	10669
QY	3061	GCGGTGGGCGAGGCGCGGAGACGCGCTGTGTGTGGGCAACCGGTCAATCGCCCTTTCGGGCG	3120
Db	10670	GCGGTGGGCGAGGCGCGGAGACGCGCTGTGTGTGGGCAACCGGTCAATCGCCCTTTCGGGCG	10729
QY	3121	GGAGCGTTTGGTCAACCACTCAACCAACGTCGGCTGGCTGGCTCGCTCGGCTCAAGGCG	3180
Db	10730	GGAGCGTTTGGTCAACCACTCAACCAACGTCGGCTGGCTGGCTCGCTCGGCTCAAGGCG	10789
QY	3181	CTCTTCGCGGATCGAGGCGGCGCGGCTCGCTGCGCTCACTTGAACGGCATGTATCGGCTCTC	3240
Db	10790	CTCTTCGCGGATCGAGGCGGCGCGGCTCGCTGCGCTCACTTGAACGGCATGTATCGGCTCTC	10849

OY	3241	GACAGAAATACC	CCCGCCCTTACG	ACCGGGGGGAG	CGGGGCTGTA	ATCCATG	CCGGCCGA	CCGGCCGG	3300
Db	10850	GACGAATTAAC	CCCGCCCTTACG	ACCGGGGGGAG	CGGGGCTGTA	ATCCATG	CCGGCCGA	CCGGCCGG	10909
OY	3301	GTCGGTCTCC	CCCGGGGAG	AGTGGGGGCG	AGCAAGTGGAG	CCGAGGTC	CAATCGA	CGGCC	3360
Db	10910	GTCGGTCTCC	CCCGGGGAG	AGTGGGGGCG	AGCAAGTGGAG	CCGAGGTC	CAATCGA	CGGCC	10965
OY	3361	GGCACGCC	CCCGGAAC	CGCGCTTAC	CTGAGAGT	CGCTGGGCGTG	CGATATGTG	AGCGATTCC	3420
Db	10970	GGCACGCC	CCCGGAAC	CGCGCTTAC	CTGAGAGT	CGCTGGGCGTG	CGATATGTG	AGCGATTCC	11029
OY	3421	CGCTCGGAG	ACGGGTTCTG	CTCGCCCA	AGTGGCGCGGTGGA	CCGGGGCGG	CGAGGAGAT	ACAACGTC	3480
Db	11030	CGCTCGGAG	ACGGGTTCTG	CTCGCCCA	AGTGGCGCGGTGGA	CCGGGGCGG	CGAGGAGAT	ACAACGTC	11089
OY	3481	GTCGCTAA	CTCGGCTTC	CGGGCGAG	CGTGAATGCA	CAAGAATTCA	TCTCCTGGAT	CGCAC	3540
Db	11090	GTCGCTAA	CTCGGCTTC	CGGGCGAG	CGTGAATGCA	CAAGAATTCA	TCTCCTGGAT	CGCAC	11149
OY	3541	GGCAGGATTG	TGGAGCTCG	GCAGAC	CGCATGTTAT	TACCGGAT	AACAGCTCG	GGCGCGG	3600
Db	11150	GGCAGGATTG	TGGAGCTCG	GCAGAC	CGCATGTTAT	TACCGGAT	AACAGCTCG	GGCGCGG	11209
OY	3601	CCGTTCTG	CGCAATCTCT	CTCTCTT	CTCGTGTG	AGATCTTC	CGGGAGAT	ATGATGCTCA	3660
Db	11210	CCGTTCTG	CGCAATCTCT	CTCTCTT	CTCGTGTG	AGATCTTC	CGGGAGAT	ATGATGCTCA	11269
OY	3661	CCGGCCGGGG	ATCCGTCCG	CTTGGAGAG	AGCTTC	CGGCTGA	TACCGGCGAG	CGGTTTC	3720
Db	11270	CCGGCCGGGG	ATCCGTCCG	CTTGGAGAG	AGCTTC	CGGCTGA	TACCGGCGAG	CGGTTTC	11329
OY	3721	ACCCCTCCCC	CAATCGCA	AGCTCC	CCGATCG	CCCGTGTG	CGCGATG	CGTTCCG	3780
Db	11330	ACCCCTCCCC	CAATCGCA	AGCTCC	CCGATCG	CCCGTGTG	CGCGATG	CGTTCCG	11389
OY	3781	GCGCAGG	CGCAGCAT	TTTGGAA	AGCTGTA	CTCA	CGCTGGGTGA	CCCGAGGTC	3840
Db	11390	GCGCAGG	CGCAGCAT	TTTGGAA	AGCTGTA	CTCA	CGCTGGGTGA	CCCGAGGTC	11449
OY	3841	CGATTTC	CAACCC	CAACGAG	CGCGCGCC	CGCTCA	CCGGGGAT	TCGGGACT	3900
Db	11450	CGATTTC	CAACCC	CAACGAG	CGCGCGCC	CGCTCA	CCGGGGAT	TCGGGACT	11509
OY	3901	CTCGCGT	CAGCTG	CGCGCGG	CGCGCGCG	CGCGCGCT	GAGGCGTT	CTCTCGTA	3960
Db	11510	CTCGCGT	CAGCTG	CGCGCGG	CGCGCGCG	CGCGCGCT	GAGGCGTT	CTCTCGTA	11569
OY	3961	GTCCTCG	AGGTGTG	CGCA	CGCCCGAA	TCAAGT	CGCGCGG	CGCTGTT	4020
Db	11570	GTCCTCG	AGGTGTG	CGCA	CGCCCGAA	TCAAGT	CGCGCGG	CGCTGTT	11629
OY	4021	CTCGG	CATG	AGCTCG	CTCATG	GCCTG	GAAGCT	GTGAGCG	4080
Db	11630	CTCGG	CATG	AGCTCG	CTCATG	GCCTG	GAAGCT	GTGAGCG	11689
OY	4081	CTGAAG	CTGTC	AGCA	CGTTCTG	TCAAGT	CGCCCA	ATATG	4140
Db	11690	CTGAAG	CTGTC	AGCA	CGTTCTG	TCAAGT	CGCCCA	ATATG	11749
OY	4141	CTGTG	ATGATCT	CTCG	CAAGCTCT	CTCTT	GGAGCG	GGTGGAGAA	4200
Db	11750	CTGTG	ATGATCT	CTCG	CAAGCTCT	CTCTT	GGAGCG	GGTGGAGAA	11809
OY	4201	GCAAG	CGTGC	AAAA	ACGATTT	GCTCAT	CGGGCG	CGATCA	4260
Db	11810	GCAAG	CGTGC	AAAA	ACGATTT	GCTCAT	CGGGCG	CGATCA	11869
OY	4261	CTATGA	4266						
Db	11870	CTATGA	11875						

```

RESULT 3
: US-09-567-969-1
: Sequence 1, Application US/09567969
: Patent No. 6355457
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goelisch, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
: FILE REFERENCE: 4-10582A
: CURRENT APPLICATION NUMBER: US/09/567,969
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 68750
: TYPE: DNA
: ORGANISM: Sorangium cellulosum
: US-09-567-969-1

```

Query Match	100.0%;	Score 4266;	DB 4;	Length 68750;
-------------	---------	-------------	-------	---------------

OY	I	TTGGGGGATTCGTTCCATCGAGCGCGCACCCGAAGA	TCCGATTGCGA	TGCTGAGACCGAGT	60
Db	7610	GTGGGGGATTCGTTCCATCGAGCGCGCACCCGAAGA	TCCGATTGCGA	TGCTGAGACCGAGT	7665
OY	61	TGCGCTTGCCCGGTGGGGGTGATCATCTGAGCGGGTT	CTTGAGACGCTCTCGAGGGCTCG	120	
Db	7670	TGCGCTTGCCCGGTGGGGGTGATCATCTGAGCGGGTT	CTTGAGACGCTCTCGAGGGCTCG	7725	
OY	121	CGCGACACCGTTCGGGCGAGTCCCCGCCGAACGCT	TGGGATTCGACGACGCTGGTTTGATCCC	180	
Db	7730	CGCGACACCGTTCGGGCGAGTCCCCGCCGAACGCT	TGGGATTCGACGACGCTGGTTTGATCCC	7785	
OY	181	GACCCGATGCCCCGGGGGAAGA	CGCGCGTTAAGCGCGATCTTTCTTGAGCGACGTAGCC	240	
Db	7790	GACCCGATGCCCCGGGGGAAGA	CGCGCGTTAAGCGCGATCTTTCTTGAGCGACGTAGCC	7845	
OY	241	TGCTTCGACGCTCTTCTTTCGGGATCTTCGCTCG	GGAAGCGCTGCGGATGGAACCTTCA	300	
Db	7850	TGCTTCGACGCTCTTCTTTCGGGATCTTCGCTCG	GGAAGCGCTGCGGATGGAACCTTCA	7905	
OY	301	CATGCACCTTCGCTGAGGTGTGCTGGGAGGGCGT	GGAAGACGCGCGATCGCTCCATCG	360	
Db	7910	CATGCACCTTCGCTGAGGTGTGCTGGGAGGGCGT	GGAAGACGCGCGATCGCTCCATCG	7965	
OY	361	GCGCTCGTTCGATACGGAACGCGGATGTTCATCG	GGGATTCGCGCTCCGAATATGAGGCC	420	
Db	7970	GCGCTCGTTCGATACGGAACGCGGATGTTCATCG	GGGATTCGCGCTCCGAATATGAGGCC	8025	
OY	421	GCGCTCGCCGAAGCGACGCGCTCCGCAAGATTCG	CGCTCATGTGCGGCTGGGGACATG	480	
Db	8030	GCGCTCGCCGAAGCGACGCGCTCCGCAAGATTCG	CGCTCATGTGCGGCTGGGGACATG	8085	
OY	481	CCGACGCTCGGAGCGGGCCGAATCTCGTATGCGCT	TCGGGCTGCGAGGGCCGCTGTCCAG	540	
Db	8090	CCGACGCTCGGAGCGGGCCGAATCTCGTATGCGCT	TCGGGCTGCGAGGGCCGCTGTCCAG	8145	
OY	541	GTGATACGCGCTATTCTGCTCGCTGCTGAGCCGTT	CACTTGACCTGTCAAGCTTTCGC	600	
Db	8150	GTGATACGCGCTATTCTGCTCGCTGCTGAGCCGTT	CACTTGACCTGTCAAGCTTTCGC	8205	
OY	601	TCCGGGGGAATGCTCCACGCGCCCTGGCTGTGGG	GGATATGCTGATGTCGCCGAGCACCC	660	
Db	8210	TCCGGGGGAATGCTCCACGCGCCCTGGCTGTGGG	GGATATGCTGATGTCGCCGAGCACCC	8265	

QY	661	TCGCGTATGGCTCTCGAAGAACCCGGGCGCTGGCCAGGGACGGATCGCTGCCTCAAGGCATTTTTCG	720
Db	8270	CTTCGTGTGGCTCTCTCGAAGAACCCGGGCGCTGGCCAGGGACGGATCGCTGCCTCAAGGCATTTTTCG	8329
QY	721	GCGAGGCGCCGATGGGATTTGCGACGAGCGGAAGGATGCGCCGTCTGATGTTCTTCAGCGGCTC	780
Db	8330	GCGAGGCGCCGATGGGATTTGCGACGAGCGGAAGGATGCGCCGTCTGATGTTCTTCAGCGGCTC	8389
QY	781	AGTGGAGCCCCCGGGAGACGGCGATTCGGATATATGGCGGTGATTTGAGAGATTCGCGATCAAT	840
Db	8390	AGTGGAGCCCCCGGGAGACGGCGATTCGGATATATGGCGGTGATTTGAGAGATTCGCGATCAAT	8449
QY	841	CACGACGGTGTGGAGAGAGCGGTCTGACCGCTGCGCAAGCGGAGCTCCCAAGAAATGTCGTG	900
Db	8450	CACGACGGTGTGGAGAGAGCGGTCTGACCGGTGCGCAAGCGGAGCTCCCAAGAAATGTCGTG	8509
QY	901	AAACGGGCGCCCTGGGCGGAGACGACAGGCTGGCGCGCGTTCCTTGATGGGTTATATGTCAGAGCAC	960
Db	8510	AAACGGGCGCCCTGGGCGGAGACGACAGGCTGGCGCGCGTTCCTTGATGGGTTATATGTCAGAGCAC	8569
QY	961	GGCACGGGCGACGACGCTTGGTGAACCCATCGAAATCCAACTCTGAAATGCGGATATACGC	1020
Db	8570	GGCACGGGCGACGACGCTTGGTGAACCCATCGAAATCCAACTCTGAAATGCGGATATACGC	8629
QY	1021	CTCGGCGGAGATATGTCGCCACGCGCCGCTGTGATCGGGTTCGGTGAAGACACACACTTGGGCAT	1080
Db	8630	CTCGGCGGAGATATGTCGCCACGCGCCGCTGTGATCGGGTTCGGTGAAGACACACACTTGGGCAT	8689
QY	1081	CCTGAGTATGAGGTGGGGGATCACTGAGGCTGTGAGAGTCTGTTCCTTCAACACAGG	1140
Db	8690	CCTGAGTATGAGGTGGGGGATCACTGAGGCTGTGAGAGTCTGTTCCTTCAACACAGG	8749
QY	1141	CAGATTCCTGGGCACCTTCACGCGCGACGGCGCTGAACCCCGGATCTCATGGGGTGAATCTT	1200
Db	8750	CAGATTCCTGGGCACCTTCACGCGCGACGGCGCTGAACCCCGGATCTCATGGGGTGAATCTT	8809
QY	1201	CGGCTGACCGCTGACGCGCGCGCCCGGACACCGTGGCGCGACCTGGAAATTAACCGCGACGGGCG	1260
Db	8810	CGGCTGACCGCTGACGCGCGCGCCCGGACACCGTGGCGCGACCTGGAAATTAACCGCGACGGGCG	8869
QY	1261	GAGGTGAGCTCGTTCCGCGATGAGCGGGACCAACCGCACGTCGTGTGCTGGAAGAGCGCGCG	1320
Db	8870	GAGGTGAGCTCGTTCCGCGATGAGCGGGACCAACCGCACGTCGTGTGCTGGAAGAGCGCGCG	8929
QY	1321	GCGGCGAGTGCACACCGCGCGCGCGCCGAGCGGACCGCGACGCTGTGATCTGTCCGCA	1380
Db	8930	GCGGCGAGTGCACACCGCGCGCGCGCCGAGCGGACCGCGACGCTGTGATCTGTCCGCA	8989
QY	1381	AGGACCGCGGTCAAGCCCTGGATATGCACAGGCGCGCGGCTGGCGGACCAATCTTGGAAACCTAC	1440
Db	8990	AGGACCGCGGTCAAGCCCTGGATATGCACAGGCGCGCGGCTGGCGGACCAATCTTGGAAACCTAC	9049
QY	1441	CCTTCGACGATGTCTGGGCGGATGTGGCGCTTCAAGTCTTGGCGACGAGCGGACGCGCATGGAG	1500
Db	9050	CCTTCGACGATGTCTGGGCGGATGTGGCGCTTCAAGTCTTGGCGACGAGCGGACGCGCATGGAG	9109
QY	1501	CACCGGCTCGCGATGGCGCGGACGTCGAGGGAGGGGCTGCGGAGGACGCTTGGAGGCTTGG	1560
Db	9110	CACCGGCTCGCGATGGCGCGGACGTCGAGGGAGGGGCTGCGGAGGACGCTTGGAGGCTTGG	9169
QY	1561	GCGCAGGGGACAGAGCTGCGCCGCGGTGGTGGTGGCGCATATGCGCCATTCTCACGCGGCAAG	1620
Db	9170	GCGCAGGGGACAGAGCTGCGCCGCGGTGGTGGTGGCGCATATGCGCCATTCTCACGCGGCAAG	9229
QY	1621	CTCGGCTTTCCTTCAACCGGACGAGGGGGCGGAGCGCTGGGCTATGGGCGGTGGCTGTAC	1680
Db	9230	CTCGGCTTTCCTTCAACCGGACGAGGGGGCGGAGCGCTGGGCTATGGGCGGTGGCTGTAC	9289
QY	1681	GATGATATGATCCGCGTTCGCGAGGCGTTCGACCTGTGCTGTGAGGCTGTTCACACAGAG	1740
Db	9290	GATGATATGATCCGCGTTCGCGAGGCGTTCGACCTGTGCTGTGAGGCTGTTCACACAGAG	9349
QY	1741	CTCGACCGGCGCTTCGCGAGGATGATGTGGCGCAACCGCGACGCTGACCGCGCGCTG	1800

Db	9350	CTCGACCGCGCGCTCCCGAGAGTATGTGGGCGGAAACCGGCGACGCTCGACCGCGCGCTGG	9409
Qy	1801	CTCGACACGACAGCGCTTTCACCCAGCGCGCGCTGTTCACCTTCGAAATATGCGCTCGCGCG	1860
Db	9410	CTCGACACGACAGCGCTTTCACCCAGCGCGCGCTGTTCACCTTCGAAATATGCGCTCGCGCG	9469
Qy	1861	CTGTGGCGGCTGTGGGGGTGTAGAACCGGAGTTGGTCCCGCGCATATGACATCGGTAGCTGG	1920
Db	9470	CTGTGGCGGCTGTGGGGGTGTAGAACCGGAGTTGGTCCCGCGCATATGACATCGGTAGCTGG	9529
Qy	1921	GTGGCTGCTCGTGGCGGCGGAGTTCCTCGCTTGGAGACCGGCGTTCCTGTGGTGGTGGG	1980
Db	9530	GTGGCTGCTCGTGGCGGCGGAGTTCCTCGCTTGGAGACCGGCGTTCCTGTGGTGGTGGG	9589
Qy	1981	CGCGGCGCGCTGATGCAAGCGCGCTCGCGCGCGCGCGGAGATGGTGTGATCGAGCGCGCG	2040
Db	9590	CGCGGCGCGCTGATGCAAGCGCGCTCGCGCGCGCGCGGAGATGGTGTGATCGAGCGCGCG	9649
Qy	2041	GAGCGCGATGTGGCTGCTGCGGCTGGCGCGCGACGACGCTCGGTGTGATCGCGCGGCT	2100
Db	9650	GAGCGCGATGTGGCTGCTGCGGCTGGCGCGCGACGACGCTCGGTGTGATCGCGCGGCT	9709
Qy	2101	AACGCTCCGACACAGGTGTGATATGCGGGGGCGGGGGAACCGGTGATATGATCGGGGG	2160
Db	9710	AACGCTCCGACACAGGTGTGATATGCGGGGGCGGGGGAACCGGTGATATGATCGGGGG	9769
Qy	2161	GCGATGGCGCGCGCGCGGGCGCGAACCAGAGCGCTTCACGCTTCGCAATGCGTTCCACTCA	2220
Db	9770	GCGATGGCGCGCGCGGGCGCGAACCAGAGCGCTTCACGCTTCGCAATGCGTTCCACTCA	9829
Qy	2221	CCGCTCATGCGCCCCGATGCTGAGAGCGTTGGGCGTGTGGCCGAGTCCGTGATCGA	2280
Db	9830	CCGCTCATGCGCCCCGATGCTGAGAGCGTTGGGCGTGTGGCCGAGTCCGTGATCGA	9889
Qy	2281	CGGCGCTCGATCGCTCGTTCAGCAATCTAGCGGGAAAGGCTTGCACAGACGAGGTGAGAC	2340
Db	9890	CGGCGCTCGATCGCTCGTTCAGCAATCTAGCGGGAAAGGCTTGCACAGACGAGGTGAGAC	9949
Qy	2341	TGCGCGGAGCTATTTGGGTGCGCACGCGCGAGAGGTGGTGGCTTCCGCGATGAGATGAGAG	2400
Db	9950	TGCGCGGAGCTATTTGGGTGCGCACGCGCGAGAGGTGGTGGCTTCCGCGATGAGATGAGAG	10009
Qy	2401	GCGCTGCACGCGGCGCGGTGCGGAGCACTTTCGTGAGGTGCGTCCGAATGACGCTGCTC	2460
Db	10010	GCGCTGCACGCGGCGCGGTGCGGAGCACTTTCGTGAGGTGCGTCCGAATGACGCTGCTC	10069
Qy	2461	GGCGTGGGTCTGTGCTGCATATCCGAGAGCGCCCGCGCGCTGCTCGCATATGTTGGCGCT	2520
Db	10070	GGCGTGGGTCTGTGCTGCATATCCGAGAGCGCCCGCGCGCTGCTCGCATATGTTGGCGCT	10129
Qy	2521	GGGCGTGCAGAGCGCGGAGACGATGCTGAGAGCGCTCGGCGGAGCTTGGGCGCGTGGAG	2580
Db	10130	GGGCGTGCAGAGCGCGGAGACGATGCTGAGAGCGCTCGGCGGAGCTTGGGCGCGTGGAG	10189
Qy	2581	CTGGTCTTCTGGGCGCGGCTTTCCTTCAGAGGGGAGGAGGGGTGCGCTGCCACGTAC	2640
Db	10190	CTGGTCTTCTGGGCGCGGCTTTCCTTCAGAGGGGAGGAGGGGTGCGCTGCCACGTAC	10249
Qy	2641	CCTTGGCAGCGCGAGCGCTACTGATATGACACCGAAAGCGACACGCGCGCGTGGCGAC	2700
Db	10250	CCTTGGCAGCGCGAGCGCTACTGATATGACACCGAAAGCGACACGCGCGCGTGGCGAC	10309
Qy	2701	CGCGGTGCTTCGAGAGCGGAGTCAAGACGAGGTGAGGAGGGGGGCGCGGTGCGCGGCG	2760
Db	10310	CGCGGTGCTTCGAGAGCGGAGTCAAGACGAGGTGAGGAGGGGGGCGCGGTGCGCGGCG	10369
Qy	2761	GACCGGCGCAGCGCTCGGCTTCGACCATTCGCGCGCGCGAGACGCGCGGAGAGGTCT	2820
Db	10370	GACCGGCGCAGCGCTCGGCTTCGACCATTCGCGCGCGCGAGACGCGCGGAGAGGTCT	10429
Qy	2821	GAGGCGCGCGCGGACGCTCCGTTCCGGCTCGAGATGATGACCGACGCGTCTTATCAC	2880

D	b	10430	GAGGCCGCCGCGGACGATCCGCTTCCGGCTCGAGATGATGAGCCAGGCGCTCTGATCAC	10485
O	y	2881	CTCGTGCCTTGGGTCACGAGCGGCGCGCCCTTGTCGTGGGCGAGGTCGAGATCGCCGTC	2940
D	b	10490	CTCGTCTTTCGGGTCACGAGCGGCGCGCCCTGGTCTGGGCGAGGTCGAGATCGCCGTC	10549
O	y	2941	GACGCGCGCGGGGCTCAGCTTCAATGATGTCCAGCTGGCGCTGGGGCATGTGCCCCGACGAC	3000
D	b	10550	GACCGCGCGGGGGCTCAGCTTCAATGATGTCCAGCTGGCGCTGGGCGATGTGCCCCGACGAC	10605
O	y	3001	CTGCGCGGAAAGCCCAACCTCCGCTGCTGCTGAGAGCGAGTGGCGCCGCGGCGCATCGATC	3060
D	b	10610	CTGCGCGGAAAGCCCAACCTCCGCTGCTGCTGAGAGCGAGTGGCGCCGCGGCGCATCGATC	10665
O	y	3061	GCCGTGGGAGGAGGCGGCTGAACGGGCTCGTGTGTGGGCAACCGGTCATCGCCCTTTCGGCG	3120
D	b	10670	GCCGTGGGAGGAGGCGGCTGAACGGGCTCGTGTGTGGGCAACCGGTCATCGCCCTTTCGGCG	10722
O	y	3121	GGAAGCTTTTGTCTACCCATCTCACACAGTGGGCTGGCTGGTGTGCTGCTCGGCTCAGGGG	3180
D	b	10730	GGAAGCTTTTGTCTACCCATCTCACACAGTGGGCTGGCTGGTGTGCTGCTCGGCTCAGGGG	10789
O	y	3181	CTCTCGGCGAATCGAGGGGCGCGGCATGCGCGTGCCTGACCTGACCGGATGTGATCGGCTC	3240
D	b	10790	CTCTCGGCGAATCGAGGGGCGCGGCATGCGCGTGCCTGACCTGACCGGATGTGATCGGCTC	10849
O	y	3241	GACAGAAATAGCCCGCCTTACGCCGCGGGGAGACGGGCTGCTGATCCATGCGGCGAGCCGGCGG	3300
D	b	10850	GACAGAAATAGCCCGCCTTACGCCGCGGGGAGACGGGCTGCTGATCCATGCGGCGAGCCGGCGG	10905
O	y	3301	GTGCGTCTCCCGCGGCTGCAGTGGGCGCAGCACGTGGAGCCGAGGTCCATCGACGAGCC	3360
D	b	10910	GTGCGTCTCCCGCGGCTGCAGTGGGCGCAGCACGTGGAGCCGAGGTCCATCGACGAGCC	10965
O	y	3361	GGACACGCCGAGAAAGCGCGCTACCTGAGTGCCTGGGGGTGGCGGATGTGAGCGCATTC	3420
D	b	10970	GGACACGCCGAGAAAGCGCGCTACCTGAGTGCCTGGGGGTGGCGGATGTGAGCGCATTC	11022
O	y	3421	CGCTCGGACCGGTTTCGTGCGCGGACGTGGCGCGCTGACCGGCGCGGAGGATGACGTC	3480
D	b	11030	CGCTCGGACCGGTTTCGTGCGCGGACGTGGCGCGCTGACCGGCGCGGAGGATGACGTC	11085
O	y	3481	GTGCTCAACTCGCTCTTGCGGCGAGCTGATCGACAGAGTTTCAATCTCTGCGATCGCAC	3540
D	b	11090	GTGCTCAACTCGCTCTTGCGGCGAGCTGATCGACAGAGTTTCAATCTCTGCGATCGCAC	11149
O	y	3541	GGCGCGTTTGTGAGAGCTCGGCGAAGCGCGACTGTTAGCGGAGTAAACAGTTCGGGCTGCGG	3600
D	b	11150	GGCGCGTTTGTGAGAGCTCGGCGAAGCGCGACTGTTAGCGGAGTAAACAGTTCGGGCTGCGG	11209
O	y	3601	CCGTTCTGTGCAATCTCTCTCTGCTGCTGTGATCTCCGGGAGATGATGTCTGAGCGG	3660
D	b	11210	CCGTTCTGTGCAATCTCTCTCTGCTGCTGTGATCTCCGGGAGATGATGTCTGAGCGG	11265
O	y	3661	CCGGCGCGGATCCGTGCGCTCTTGTGAGAGCTCTCGACTGATCGGCGAGCGCTGTTC	3720
D	b	11270	CCGGCGCGGATCCGTGCGCTCTTGTGAGAGCTCTCGACTGATCGGCGAGCGCTGTTC	11329
O	y	3721	ACCCCTCCCCCCTATCCGCGAAGCTCCCGGATCGCCCGTGTGGCGGATGCGGAGCATG	3780
D	b	11330	ACCCCTCCCCCCTATCCGCGAAGCTCCCGGATCGCCCGTGTGGCGGATGCGGAGCATG	11389
O	y	3781	GCGCAGGCGAGCATCTTGTGGAAGCTCGTACTACGCTGGGTGACC CGGAGGTCAGATC	3840
D	b	11390	GCGCAGGCGAGCATCTTGTGGAAGCTCGTACTACGCTGGGTGACC CGGAGGTCAGATC	11449
O	y	3841	CGTATTCACACCCAGCGAGGCGCGGCGCTTCACCGGGGATCGGAGACTGCTGACAGG	3900
D	b	11450	CGTATTCACACCCAGCGAGGCGCGGCGCTTCACCGGGGATCGGAGACTGCTGACAGG	11509
O	y	3901	CTGCGCTCAGATCGCGCGCGCGCGCGCGCGCGCGCTGGAAGCGTTCCTCCGTATCGCAG	3960
D	b	11510	CTGCGCTCAGATCGCGCGCGCGCGCGCGCGCGCGCTGGAAGCGTTCCTCCGTATCGCAG	11569

QY 3961 GTCTCGAGGTGCTGCGCAGACCCCGAATCAAGGTGCGCGAGCGCTGTTCAACCGC 4020
|
|
|
Db 11570 GTCTCGAGGTGCTGCGCAGACCCCGAATCAAGGTGCGCGAGCGCTGTTCAACCGC 11629
|
|
|
QY 4021 CTGGGCACTGCACTGCTCATAGGCGGTGAGCTGCGCAATGTAATCGAGGAGGAGCTCAAG 4080
|
|
|
Db 11630 CTGGGCACTGCACTGCTCATAGGCGGTGAGCTGCGCAATGTAATCGAGGAGGAGCTCAAG 11689
|
|
|
QY 4081 CTGAAGCTGTCAGACGAGTCTGTCTGTCACAGTCCCGCAATATCCGCTTGTTGAGCCAAAC 4140
|
|
|
Db 11690 CTGAAGCTGTCAGACGAGTCTGTCTGTCACAGTCCCGCAATATCCGCTTGTTGAGCCAAAC 11749
|
|
|
QY 4141 CTGTTGAGTCTCTGCGCAGCTCTCTCTCTGAGCGGTGCGCGAGAACTTAACG 4200
|
|
|
Db 11750 CTGTTGAGTCTCTGCGCAGCTCTCTCTCTGAGCGGTGCGCGAGAACTTAACG 11809
|
|
|
QY 4201 GCAAGGCTGCAAAACGACTTCTGTCATCGGCGCGAGATCAAGACTGGAATCATTTGCC 4260
|
|
|
Db 11810 GCAAGGCTGCAAAACGACTTCTGTCATCGGCGCGAGATCAAGACTGGAATCATTTGCC 11869
|
|
|
QY 4261 CTATGA 4266
|
|
|
Db 11870 CTATGA 11875

RESULT 4

US-09-568-480-1

Sequence 1, Application US/09568480

Patent No. 6355458

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James

APPLICANT: Molnar, Istvan

APPLICANT: Zickler, Ross

APPLICANT: Cyr, Devon

APPLICANT: Goerlach, Joern

TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES

FILE REFERENCE: 4-30582A

CURRENT APPLICATION NUMBER: US/09/568,480

CURRENT FILING DATE: 2000-05-10

PRIOR APPLICATION NUMBER: 09/335,409

PRIOR FILING DATE: 1999-06-17

NUMBER OF SEQ ID NOS: 30

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 68750

TYPE: DNA

ORGANISM: Sorangium cellulosum

US-09-568-480-1

Query Match 100.0%; Score 4266; DB 4; Length 68750;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGGCGGATGTCCTCCATGAGCGCGGACCCGAAAGATCCGATTGGGATGTGCGAGCGAGT 60
|
|
|
Db 7610 GTGGCGGATGTCCTCCATGAGCGCGGACCCGAAAGATCCGATTGGGATGTGCGAGCGAGT 7669
|
|
|
QY 61 TGCCGTGCGCGGCGGTGATGATGTGAGCGCGGTTCTGAGCGCTCTGAGGCGCTCG 120
|
|
|
Db 7670 TGCCGTGCGCGGCGGTGATGATGTGAGCGCGGTTCTGAGCGCTCTGAGGCGCTCG 7729
|
|
|
QY 121 CGCGACACCGTGGGCGAGTCCCGCGCGAAGCGTGGAGTGAAGCAGCGTGGTTGATCC 180
|
|
|
Db 7730 CGCGACACCGTGGGCGAGTCCCGCGCGAAGCGTGGAGTGAAGCAGCGTGGTTGATCC 7789
|
|
|
QY 181 GACCCGATGCGCGCGGAGAGCGCGGTAAAGCGCGGATCTTTCTTGAAGCAGCTAGCC 240
|
|
|
Db 7790 GACCCGATGCGCGCGGAGAGCGCGGTAAAGCGCGGATCTTTCTTGAAGCAGCTAGCC 7849
|
|
|
QY 241 TGCTTGAAGCGCTCTTCTTGGGATGTCGCGCTCGCGAAGCGCTGCGATGAGACCTTCA 300
|
|
|

Db 7850 TGCTTGAAGCGCTCTTCTTGGGATGTCGCGCTCGCGAAGCGCTGCGATGAGACCTTCA 7909
|
|
|
QY 301 CATGCACTCTTGTGAGAGGTGTGCTGGAGGCGCTGGAAGACGCCCGCATTCGTCATG 360
|
|
|
Db 7910 CATGCACTCTTGTGAGAGGTGTGCTGGAGGCGCTGGAAGACGCCCGCATTCGTCATG 7969
|
|
|
QY 361 GCGCTGTGCTGATCGGAAACCGGAGTGTTCATCGGAGTGGGCCGCTCCGAATATAGAGCC 420
|
|
|
Db 7970 GCGCTGTGCTGATCGGAAACCGGAGTGTTCATCGGAGTGGGCCGCTCCGAATATAGAGCC 8029
|
|
|
QY 421 GCGCTGTGCTGATCGGAGCGGCTGCGAGAGTGAAGCTCATGAGCGGCTGGGAGAGT 480
|
|
|
Db 8030 GCGCTGTGCTGATCGGAGCGGCTGCGAGAGTGAAGCTCATGAGCGGCTGGGAGAGT 8089
|
|
|
QY 481 CCCAGGCTGAGAGCGGCGGAAATCTGTATGCTCTGGGCTGGAGAGGCGGTGTTCGCG 540
|
|
|
Db 8090 CCCAGGCTGAGAGCGGCGGAAATCTGTATGCTCTGGGCTGGAGAGGCGGTGTTCGCG 8149
|
|
|
QY 541 GTGGATATCGGCTCAATTCGTCGCTGAGGCGCTTCATCTGAGCTGTGAGAGCTTGGCG 600
|
|
|
Db 8150 GTGGATATCGGCTCAATTCGTCGCTGAGGCGCTTCATCTGAGCTGTGAGAGCTTGGCG 8209
|
|
|
QY 601 TCCGAGGAAATGCTCAAGCGGCTGCTGAGGAGTATCGCTGATGTTGTGCGGAGCACC 660
|
|
|
Db 8210 TCCGAGGAAATGCTCAAGCGGCTGCTGAGGAGTATCGCTGATGTTGTGCGGAGCACC 8269
|
|
|
QY 661 CTGCTGTGCTCTGCAAGACCCGCGCTGCGCAAGGAGCGGTGCTGCAAGGATTTTCG 720
|
|
|
Db 8270 CTGCTGTGCTCTGCAAGACCCGCGCTGCGCAAGGAGCGGTGCTGCAAGGATTTTCG 8329
|
|
|
QY 721 GCGGAGGCGGATGAGGTTTCGAGCAGAGGAGAAAGGCTGCGCTGCTCAAGCGGCTG 780
|
|
|
Db 8330 GCGGAGGCGGATGAGGTTTCGAGCAGAGGAGAAAGGCTGCGCTGCTCAAGCGGCTG 8389
|
|
|
QY 781 AGTGAAGCCCGCGCGAGCGGATGAGATTTGCGGAGTATGAGAGATCGCGATCAAT 840
|
|
|
Db 8390 AGTGAAGCCCGCGCGAGCGGATGAGATTTGCGGAGTATGAGAGATCGCGATCAAT 8449
|
|
|
QY 841 CAGGACGCTGCGAGCAGCGGCTGATCGCTGCGCAAGGAGCTCCCAAGAAATCTGCTG 900
|
|
|
Db 8450 CAGGACGCTGCGAGCAGCGGCTGATCGCTGCGCAAGGAGCTCCCAAGAAATCTGCTG 8509
|
|
|
QY 901 AAACGGGCGCTGAGCGAGCGGCTGCGCGCTCTGCTGAGGATTTGCTGAGGAGCAGC 960
|
|
|
Db 8510 AAACGGGCGCTGAGCGAGCGGCTGCGCGCTCTGCTGAGGATTTGCTGAGGAGCAGC 8569
|
|
|
QY 961 GGCACGGGCAAGACGCTTGTGTAACCCCATGAATCCAGCTTGAATGCGGTATACGGC 1020
|
|
|
Db 8570 GGCACGGGCAAGACGCTTGTGTAACCCCATGAATCCAGCTTGAATGCGGTATACGGC 8629
|
|
|
QY 1021 CTGCGGCGGATGTCGCAAGCGGCTGCTGATCGGCTGCTGTAAGAACCACTTTGGCGAT 1080
|
|
|
Db 8630 CTGCGGCGGATGTCGCAAGCGGCTGCTGATCGGCTGCTGTAAGAACCACTTTGGCGAT 8689
|
|
|
QY 1081 CCGAGATGATGCTGAGGAGTCACTGAGCTGTAAGGCTGCTGCTTCTTGAAGCAGCGG 1140
|
|
|
Db 8690 CCGAGATGATGCTGAGGAGTCACTGAGCTGTAAGGCTGCTGCTTCTTGAAGCAGCGG 8749
|
|
|
QY 1141 CAGATTCTGCGACCTTCAAGCGGCGAGGCTGAACCCCGGATCTCATGAGGATGATCTT 1200
|
|
|
Db 8750 CAGATTCTGCGACCTTCAAGCGGCGAGGCTGAACCCCGGATCTCATGAGGATGATCTT 8809
|
|
|
QY 1201 CCGCTGAACCTGACGCGCGCGCGAGACCGTGTGCGGATGTAAGTACGCGCGAGCGGCG 1260
|
|
|
Db 8810 CCGCTGAACCTGACGCGCGCGCGAGACCGTGTGCGGATGTAAGTACGCGCGAGCGGCG 8869
|
|
|
QY 1261 GGGGTGAGCTGCTTGGGATGAGGAGGAGCAACCGCAGCTGAGTGTGTAAGAGGCGCG 1320
|
|
|
Db 8870 GGGGTGAGCTGCTTGGGATGAGGAGGAGCAACCGCAGCTGAGTGTGTAAGAGGCGCG 8929
|
|
|
QY 1321 GCGGCGAGCTGACAGCGCGCGCGCGAGGAGCGGCAAGAGCTGCTGCTGCTGCGCA 1380
|
|
|
Db 8930 GCGGCGAGCTGACAGCGCGCGCGCGAGGAGCGGCAAGAGCTGCTGCTGCTGCGCA 8989

QY 1381 AGGACCGCGTCAAGCCCTTGATGACACAGCGCGCGCGCTGCGGACCATCTGAGACTTAC 1440
DB 8990 AGGACCGCGTCAAGCCCTTGATGACACAGCGCGCGCGCTGCGGACCATCTGAGACTTAC 9049
QY 1441 CTTTGCGATGCTCTGGGCGCATGTTGGCGTTCACTTGGCGACACAGCGCGCGCATATGAG 1500
DB 9050 CTTTGCGATGCTCTGGGCGCATGTTGGCGTTCACTTGGCGACACAGCGCGCGCATATGAG 9109
QY 1501 CACCGGCTCGCGGTGCGCGCGCGCATGTCAGAGGAGGAGGCTGCGGCGACCCCTGAGCGCTGCG 1560
DB 9110 CACCGGCTCGCGGTGCGCGCGCGCATGTCAGAGGAGGAGGCTGCGGCGACCCCTGAGCGCTGCG 9169
QY 1551 GCGCAGGAGACAGACGTCGCGCGGTGCGGTGCGAGTATGCGGATTCCTCAAGCGGACAG 1620
DB 9170 GCGCAGGAGACAGACGTCGCGCGGTGCGGTGCGAGTATGCGGATTCCTCAAGCGGACAG 9229
QY 1621 CTCGCTTTCTTTTCAACGCGACAGGAGGCGCGACAGCTGGGCAATGGGCGGTGGCTTAC 1680
DB 9230 CTCGCTTTCTTTTCAACGCGACAGGAGGCGCGACAGCTGGGCAATGGGCGGTGGCTTAC 9289
QY 1681 GATGATGATGCTCGCGGTTCGCGAGGCGTTGACCTGTGCTGAGGCTGTTCAACAGAGAG 1740
DB 9290 GATGATGATGCTCGCGGTTCGCGAGGCGTTGACCTGTGCTGAGGCTGTTCAACAGAGAG 9349
QY 1741 CTCGACCGGCGCTCGCGAGGTGATGTTGGCGCGAACCGGCGAGCTGACCGCGCTG 1800
DB 9350 CTCGACCGGCGCTCGCGAGGTGATGTTGGCGCGAACCGGCGAGCTGACCGCGCTG 9409
QY 1801 CTCGACCGAGACGCTTCAACCAAGCGCGCGCTGTTCACTTGAATATGCGTCCGCGCG 1860
DB 9410 CTCGACCGAGACGCTTCAACCAAGCGCGCGCTGTTCACTTGAATATGCGTCCGCGCG 9469
QY 1861 CTGTCGCGGTGCTGAGGCTGTAAGCGCGAGGTTGTCGCGCGCATATGAGATGCGTGAAGCTG 1920
DB 9470 CTGTCGCGGTGCTGAGGCTGTAAGCGCGAGGTTGTCGCGCGCATATGAGATGCGTGAAGCTG 9529
QY 1921 GTGCTGCTGCTGCTGCGCGCGCGCTGTTCTGCTTGAAGACGCGGTGTTCTGCTGGCTGCG 1980
DB 9530 GTGCTGCTGCTGCTGCGCGCGCGCTGTTCTGCTTGAAGACGCGGTGTTCTGCTGGCTGCG 9589
QY 1981 GCGCGGCGCTGATGACAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
DB 9590 GCGCGGCGCTGATGACAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9649
QY 2041 GAGGCGCATGTCGCTGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
DB 9650 GAGGCGCATGTCGCTGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9709
QY 2101 AACGCTCGGACACAGTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2160
DB 9710 AACGCTCGGACACAGTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9769
QY 2161 GCGATGCG 2220
DB 9770 GCGATGCG 9829
QY 2221 CCGCTCATGCG 2280
DB 9830 CCGCTCATGCG 9889
QY 2281 CCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
DB 9890 CCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9949
QY 2341 TCGCGCGCTATGCGGTGCG 2400
DB 9950 TCGCGCGCTATGCGGTGCG 10009
QY 2401 GCGCTGACAGCG 2460
DB 10010 GCGCTGACAGCG 10069

QY 2461 GCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2520
DB 10070 GCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10129
QY 2521 GCGCGTGAACAGCG 2580
DB 10130 GCGCGTGAACAGCG 10189
QY 2581 CTGCTCTCTGCG 2640
DB 10190 CTGCTCTCTGCG 10249
QY 2641 CTTTGGCAGCGCGGAGCTGATGATGACACAGAAAGCTGACGACCGCGCGCTGCGGAC 2700
DB 10250 CTTTGGCAGCGCGGAGCTGATGATGACACAGAAAGCTGACGACCGCGCGCTGCGGAC 10309
QY 2701 CGCGTGTCTCGCGGAGCGGCTGACAGAGGTTGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 2760
DB 10310 CGCGTGTCTCGCGGAGCGGCTGACAGAGGTTGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 10369
QY 2761 GACCGGCGCAGCGCTCGGCTGACACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2820
DB 10370 GACCGGCGCAGCGCTCGGCTGACACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10429
QY 2821 GAGCGCGCGCGGACCGCTCGCTTCCGCTGCGAGATGATGACAGCGCGCTGCTGATCAC 2880
DB 10430 GAGCGCGCGCGGACCGCTCGCTTCCGCTGCGAGATGATGACAGCGCGCTGCTGATCAC 10489
QY 2881 CTGCTGCTTCCGCTGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2940
DB 10490 CTGCTGCTTCCGCTGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10549
QY 2941 GACCGGCGCGGCTCAGCTTCAATGATGATGACAGCTCGCGCTGCGGAGTGGCGGACGAC 3000
DB 10550 GACCGGCGCGGCTCAGCTTCAATGATGATGACAGCTCGCGCTGCGGAGTGGCGGACGAC 10609
QY 3001 CTGCGGAGAAAGCCCAACCTTCGCTGCTGCTGCGAGGCGAGTGGCGCGGCGCATGCTC 3060
DB 10610 CTGCGGAGAAAGCCCAACCTTCGCTGCTGCTGCGAGGCGAGTGGCGCGGCGCATGCTC 10669
QY 3061 GCGCTGGCGAGGCGGTGACAGCGCTGCTGCTGCGGCGCAACCGGTATGCGCTTTCGCGCG 3120
DB 10670 GCGCTGGCGAGGCGGTGACAGCGCTGCTGCTGCGGCGCAACCGGTATGCGCTTTCGCGCG 10729
QY 3121 GAGGCTTGTGCTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGT 3180
DB 10730 GAGGCTTGTGCTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 10789
QY 3181 CTCTCGCGGATGAGGCGCGCGCGCATGCGCTGCGGCTGCTGACGCGCATGCTGACCGGCTC 3240
DB 10790 CTCTCGCGGATGAGGCGCGCGCGCATGCGCTGCGGCTGCTGACGCGCATGCTGACCGGCTC 10849
QY 3241 GACGATATGCGCGCTTCAAGCGCGGAGCGGCTGCTGATTCATGCGCGCACCGCGCGCG 3300
DB 10850 GACGATATGCGCGCTTCAAGCGCGGAGCGGCTGCTGATTCATGCGCGCACCGCGCGCG 10909
QY 3301 GTGCGTCTCGCGCGGTGACAGTGGCGGACAGCGGAGGCGGAGTGCATGCGACGCGCG 3360
DB 10910 GTGCGTCTCGCGCGGTGACAGTGGCGGACAGCGGAGGCGGAGTGCATGCGACGCGCG 10969
QY 3361 GGCAGCGCGCGGAAAGCGGCTTACCTGAGTGCCTGAGCGGCTGCGGATATGAGGAGCTTCC 3420
DB 10970 GGCAGCGCGCGGAAAGCGGCTTACCTGAGTGCCTGAGCGGCTGCGGATATGAGGAGCTTCC 11029
QY 3421 CGCTCGGACCGGCTTCTGCGCGAGCGTGCAGCGGCGGCGGCGGCGGAGGAGTACGCTC 3480
DB 11030 CGCTCGGACCGGCTTCTGCGCGAGCGTGCAGCGGCGGCGGCGGAGGAGTACGCTC 11089
QY 3481 GTGCTCAATGCTGCTGCGGCGGAGCTGATGACAAAGTTCATCTCTGCGATGCGAC 3540
DB 11090 GTGCTCAATGCTGCTGCGGCGGAGCTGATGACAAAGTTCATCTCTGCGATGCGAC 11149
QY 3541 GCGCGGTTTGTGAGCTCGGACGCGCATGTTTACGCGGATTAACAGCTCGGCGTGGCG 3600

Db	11550	GGCCCGGTTTGGAGCTCGGCAAGCGGACCTGTTAGCCGATTAACAGCTCGGGCTGCGG	11209
Qy	3601	CCGTTCTGGGCAATCTCTCTCTCTCGCTGGTGGATCTCCGGGGATGATGCTCGAGCGG	3660
Db	11210	CCGTTCTGGGCAATCTCTCTCTCTCGCTGGTGGATCTCCGGGGATGATGCTCGAGCGG	11265
Qy	3661	CCGGCGGGTTCGGTGGCTCTTGAGAGAGCTCTCGGCTGATCCGCGAGCGCGTGTTC	3720
Db	11270	CCGGCGGGTTCGGTGGCTCTTGAGAGAGCTCTCGGCTGATCCGCGAGCGCGTGTTC	11329
Qy	3721	ACCCCTCCCCCATCCGAGAGCTCCGATCCCGCTGTGCGCGATGCTTCCGAGCATG	3780
Db	11330	ACCCCTCCCCCATCCGAGAGCTCCGATCCCGCTGTGCGCGATGCTTCCGAGCATG	11389
Qy	3781	GGCGAGCGGAGCATCTTGGGAAAGCTGTACTACGCTGGGTTACCCGGAGGTCAGATC	3840
Db	11390	GGCGAGCGGAGCATCTTGGGAAAGCTGTACTACGCTGGGTTACCCGGAGGTCAGATC	11449
Qy	3841	GGTATTCCAAACCCAGCAGGCGCGCGCCGCTCCACCGGGATCGGAACTCTGTCAGAG	3900
Db	11450	GGTATTCCAAACCCAGCAGGCGCGCGCCGCTCCACCGGGATCGGAACTCTGTCAGAG	11509
Qy	3901	CTCGGCTCAGCTCGGCGCGCGCGCGCGCGCGCGCTGAGAGGCTTCTCCGTAGCAG	3960
Db	11510	CTCGGCTCAGCTCGGCGCGCGCGCGCGCGCGCGCTGAGAGGCTTCTCCGTAGCAG	11569
Qy	3961	GTCTCGAGGTTGCTGCCACGCCCAAAATCAAGTCCGCGCGAGCGCTTCAACCGC	4020
Db	11570	GTCTCGAGGTTGCTGCCACGCCCAAAATCAAGTCCGCGCGAGCGCTTCAACCGC	11629
Qy	4021	CTCGGCAATGCACTGCTCATAGCCGCTGGAAGCTGCGCAATGTTTCAAGGAGCCTCAAG	4080
Db	11630	CTCGGCAATGCACTGCTCATAGCCGCTGGAAGCTGCGCAATGTTTCAAGGAGCCTCAAG	11689
Qy	4081	CTGAAGCTGTGCAAGCAGTCTCTGTCAAGTCCGCCCAATATCCCTTGTGGCCCAAAAC	4140
Db	11690	CTGAAGCTGTGCAAGCAGTCTCTGTCAAGTCCGCCCAATATCCCTTGTGGCCCAAAAC	11749
Qy	4141	CTGTTGATGCTCTCGCCACAGCTCTCTCTTGGAGCGGGTGGCGGAGAACTTAGCG	4200
Db	11750	CTGTTGATGCTCTCGCCACAGCTCTCTCTTGGAGCGGGTGGCGGAGAACTTAGCG	11809
Qy	4201	GCAGGCGTGCAGAAACGACTTCGCTCATCGGGGCGAGTCAACATCGGAAATCATTTGCC	4260
Db	11810	GCAGGCGTGCAGAAACGACTTCGCTCATCGGGGCGAGTCAACATCGGAAATCATTTGCC	11869
Qy	4261	CTATGCA 4266	
Db	11870	CTATGCA 11875	
RESULT 5			
DS-09-568-486-1			
Sequence 1, Application US/09568486			
Patent No. 6355459			
GENERAL INFORMATION:			
APPLICANT: Schupp, Thomas			
APPLICANT: Ligon, James			
APPLICANT: Molnar, Istvan			
APPLICANT: Zirkle, Ross			
APPLICANT: Cyr, Devon			
APPLICANT: Goerlach, Joern			
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES			
FILE REFERENCE: 4-30582A			
CURRENT APPLICATION NUMBER: US/09/568,486			
CURRENT FILING DATE: 2000-05-10			
PRIOR APPLICATION NUMBER: 09/335,409			
PRIOR FILING DATE: 1999-06-17			
NUMBER OF SEQ ID NOS: 30			
SOFTWARE: PatentIn Ver. 2.0			
SEQ ID NO 1			
LENGTH: 68750			

Query Match	100.0%	Score 4266	DB 4	Length 68750
Best Local Similarity	100.0%	Pident. No. 0	Mismatches 0	Indels 0
Matches 4266	Conservative 0	Matches 0	Indels 0	Gaps 0
US-09-568-486-1				
TYPE: DNA				
ORGANISM: Sorangium cellulosum				
1 GTGCGCATCGTCCCATCTGACGCGCGACCGAAGATCCGATTGCGATCTGTCGAGCGAGT	60			
7610 GTGCGCATCGTCCCATCTGACGCGCGACCGAAGATCCGATTGCGATCTGTCGAGCGAGT	7669			
61 TGCCGTCGCCCGGTGCGGTGATCTGATGAGCGGATTCTGACGCTCTCTGAGGCTCG	120			
7670 TGCCGTCGCCCGGTGCGGTGATCTGATGAGCGGATTCTGACGCTCTCTGAGGCTCG	7729			
121 CGGCAKACCGTCCGCGGAGTCCCGCGGAAACGCTGGGATGACAGCGGTGTTGATCCC	180			
7730 CGGCAKACCGTCCGCGGAGTCCCGCGGAAACGCTGGGATGACAGCGGTGTTGATCCC	7789			
181 GACCCCGATGCCCCGGGGAGACGCGCGTTACGCGCATCTTCTGAGCGAGTACG	240			
7790 GACCCCGATGCCCCGGGGAGACGCGCGTTACGCGCATCTTCTGAGCGAGTACG	7849			
241 TGCTTCGACGCTCTCTTCTGCGATCTGCGATCTGCGAACGCTGCGGATGAGACCGTCA	300			
7850 TGCTTCGACGCTCTCTTCTGCGATCTGCGATCTGCGAACGCTGCGGATGAGACCGTCA	7909			
301 CATGACCTCTGCTGAGAGTGTGCTGGGAGCGCTGAGAGACGCGCGATCGCTCATCG	360			
7910 CATGACCTCTGCTGAGAGTGTGCTGGGAGCGCGCTGAGAGACGCGCGATCGCTCATCG	7969			
361 GCCTCTCGTCCGATCGAAGACGCGAGTGTTCATCGGATCGGCGCTCGGAATATGAGCC	420			
7970 GCCTCTCGTCCGATCGAAGACGCGAGTGTTCATCGGATCGGCGCTCGGAATATGAGCC	8029			
421 GCGCTGCGCAAGGAGCGCGCTCGCGAGATCGACGCTCATGCGGCGTGGGACGATG	480			
8030 GCGCTGCGCAAGGAGCGCGCTCGCGAGATCGACGCTCATGCGGCGTGGGACGATG	8089			
481 CCCAGCGTCGAGGCGGCGCGAATCTCGTATGCCCTCGGCGTGCAGGCGCGTGTCTCG	540			
8090 CCCAGCGTCGAGGCGGCGCGAATCTCGTATGCCCTCGGCGTGCAGGCGCGTGTCTCG	8149			
541 GTGATATGCGCTATTTGCTCTGCTGCTGTGGCCCTTATCTGGCTCTCAGAGTTTCCG	600			
8150 GTGATATGCGCTATTTGCTCTGCTGCTGTGGCCCTTATCTGGCTCTCAGAGTTTCCG	8209			
601 TCCGAGGATCTCTCAAGCGCTCTGCTGCTGAGGATCTGCTGATGTTGTGCGCCAGAC	660			
8210 TCCGAGGATCTCTCAAGCGCTCTGCTGCTGAGGATCTGCTGATGTTGTGCGCGAGAC	8269			
661 CTGCTGTGCTCTTGAAGACCTCGGCGCTGCGCCAGGACGCTGCTGCAAGGCAATTTTC	720			
8270 CTGCTGTGCTCTTGAAGACCTCGGCGCTGCGCCAGGACGCTGCTGCAAGGCAATTTTC	8329			
721 GCGAGGCGCGATGCGTCTGCAAGGAGGAGGCGCCCGTCTGCTCTCAAGCGCTC	780			
8330 GCGAGGCGCGATGCGTCTGCAAGGAGGAGGCGCCCGTCTGCTCTCAAGCGCTC	8389			
781 AGTGAAGCGCGCGAGCGCGATCGATATTTGGCGGTGATTGAGGATCCGATCAAT	840			
8390 AGTGAAGCGCGCGAGCGCGATCGATATTTGGCGGTGATTGAGGATCCGATCAAT	8449			
841 CACGACGCTGGAACGCGGTCTGACCGTGCAGAACGCGAGCTCCCAAGAAATGTCTG	900			
8450 CACGACGCTGGAACGCGGTCTGACCGTGCAGAACGCGAGCTCCCAAGAAATGTCTG	8509			
901 AAACGGGCGCTGCGGACGCAAGCTGCGCGCTCTTCTGCTGAGTTATGTCGAGGACAC	960			
8510 AAACGGGCGCTGCGGACGCAAGCTGCGCGCTCTTCTGCTGAGTTATGTCGAGGACAC	8569			
961 GGCAGGCGCAGACGCTTGTGACCCCATCGAAATCCAGCTCTGATGCGGTATACGC	1020			

Db 8570 GGCACGGGACAGACGCTGTGTGACCCCAATCGAAATCCAAAGCTCTGAAGCGGTAAATACGGC 8629
Qy 1021 CTCGGGACAGATGTTCGACGCGCGCTGCTGATCGGGTGGGTGAAGACCAACCTTGACCAT 1080
Db 8630 CTCGGGACAGATGTTCGACGCGCGCTGCTGATCGGGTGGGTGAAGACCAACCTTGACCAT 8689
Qy 1081 CTTGATATGCGTCGGGGATCACTGGGGTGTGTGAAGTGTCTTGTCTTCCCTTCAGACGGG 1140
Db 8690 CTTGATATGCGTCGGGGATCACTGGGGTGTGTGAAGTGTCTTGTCTTCCCTTCAGACGGG 8749
Qy 1141 CAGATTCTGGGACCTTCAAGCGCGACGGCGCTGAACCCCGGATTTCAATGGGTGATCTT 1200
Db 8750 CAGATTCTGGGACCTTCAAGCGCGACGGCGCTGAACCCCGGATTTCAATGGGTGATCTT 8809
Qy 1201 CGGGTGAACGTCAGCGGGGCGCGGACACGTCGGCGCGA CTGGAAATACCGCGGACGGGCG 1260
Db 8810 CGGGTGAACGTCAGCGGGGCGCGGACACGTCGGCGCGA CTGGAAATACCGCGGACGGGCG 8869
Qy 1261 GGGGTGAACGTCGTCGATGACGGGACCAACGCGACGTCGTCGTCGTCGTCGTCGTCGTCGTC 1320
Db 8870 GGGGTGAACGTCGTCGATGACGGGACCAACGCGACGTCGTCGTCGTCGTCGTCGTCGTCGTC 8929
Qy 1321 GCGGCGACGTGACACCGCGCGCGCGCGGACCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTC 1380
Db 8930 GCGGCGACGTGACACCGCGCGCGCGGACCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTC 8989
Qy 1381 AGGACCGGTCAGCGCTGGATGACAGGCGGGCGGGCTGCGGACCATCTGGAGCCTAC 1440
Db 8990 AGGACCGGTCAGCGCTGGATGACAGGCGGGCGGGCTGCGGACCATCTGGAGCCTAC 9049
Qy 1441 CTTTCGACGTCGTCGGGATGTGGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1500
Db 9050 CTTTCGACGTCGTCGGGATGTGGCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9105
Qy 1501 CACCGGCTTCGGGGTGGCGGACGTCGACGAGGGGGCTGCGGACGCTTGGACGCTGCG 1560
Db 9110 CACCGGCTTCGGGGTGGCGGACGTCGACGAGGGGGCTGCGGACGCTTGGACGCTGCG 9169
Qy 1561 GCGGACGGGACAGAGTCGCGCGGGTGGCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1620
Db 9170 GCGGACGGGACAGAGTCGCGCGGGTGGCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9229
Qy 1621 CTCGCTTCTTTCACCGGACAGGGGGCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1680
Db 9230 CTCGCTTCTTTCACCGGACAGGGGGCGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9289
Qy 1681 GATGATATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1740
Db 9290 GATGATATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9349
Qy 1741 CTCGACCGGCGCGCTCGGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1800
Db 9350 CTCGACCGGCGCGCTCGGAGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9409
Qy 1801 CTCGACCGGACGCTTACCGACGCGCGCGCTGTCACCTTCGATATGTCGTCGTCGTCGTC 1860
Db 9410 CTCGACCGGACGCTTACCGACGCGCGCGCTGTCACCTTCGATATGTCGTCGTCGTCGTC 9469
Qy 1861 CTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1920
Db 9470 CTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9529
Qy 1921 GTC 1980
Db 9530 GTC 9589
Qy 1981 GCGGCGCGCTGATGACAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
Db 9590 GCGGCGCGCTGATGACAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9649
Qy 2041 GAGGCGGATGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2100

Db 9650 GAGGCGGATGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9709
Qy 2101 AACGCTCCGACAGAGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2160
Db 9710 AACGCTCCGACAGAGTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9769
Qy 2161 GCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2220
Db 9770 GCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9829
Qy 2221 CCGCTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2280
Db 9830 CCGCTCATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9889
Qy 2281 CCGGCGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2340
Db 9890 CCGGCGTCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 9949
Qy 2341 TCGCCGCGCTATTTGGGTGCGCCACGCGGAGAGGTGCGCTTCGCGGATGAGTCGTCGTCGTCGTC 2400
Db 9950 TCGCCGCGCTATTTGGGTGCGCCACGCGGAGAGGTGCGCTTCGCGGATGAGTCGTCGTCGTCGTC 10009
Qy 2401 GCGCTGACG 2460
Db 10010 GCGCTGACG 10069
Qy 2461 GCGCTGTC 2520
Db 10070 GCGCTGTC 10129
Qy 2521 GGGGTCGACGAGCG 2580
Db 10130 GGGGTCGACGAGCG 10189
Qy 2581 CTCGTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2640
Db 10190 CTCGTCCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10249
Qy 2641 CTTTCGACGCGGACGCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2700
Db 10250 CTTTCGACGCGGACGCTACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10309
Qy 2701 GCGGTC 2760
Db 10310 GCGGTC 10369
Qy 2761 GACCGGCGGACGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2820
Db 10370 GACCGGCGGACGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10429
Qy 2821 GAGGCGCGCGGACGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2880
Db 10430 GAGGCGCGCGGACGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10489
Qy 2881 CTCGTC 2940
Db 10490 CTCGTC 10549
Qy 2941 GACCGGCGGCGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3000
Db 10550 GACCGGCGGCGCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10609
Qy 3001 CTCGCGGGAAGCCCAACCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3060
Db 10610 CTCGCGGGAAGCCCAACCTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10669
Qy 3061 GCGGTC 3120
Db 10670 GCGGTC 10729
Qy 3121 GAGGCGCTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 3180
Db 10730 GAGGCGCTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 10789

OY	3181	CTCTCGGGAATCGAGGGGGCCGCAATCCGCTGCGATCTGAACGAGATGTTAGCGGCTC	3240
Db	10790	CTCTCGGGAATCGAGGGGGCCGCAATCCGCTGCGATCTGAACGAGATGTTAGCGGCTC	10849
OY	3241	GACAGAAATAGCCCGCTTTCAGCCGGGGGAGCGGGTGTGATTCATGCGCGCACCGGCGGG	3300
Db	10850	GACAGAAATAGCCCGCTTTCAGCCGGGGGAGCGGGTGTGATTCATGCGCGCACCGGCGGG	10908
OY	3391	GTCGGTCTCGCGGGGTGAGTAGTGGGGCAGACGATGGGAACCGAGGTCCATGTGACGGGCC	3360
Db	10910	GTCGGTCTCGCGGGGTGAGTAGTGGGGCAGACGATGGGAACCGAGGTCCATGTGACGGGCC	10966
OY	3361	GGCAGCGCCCGAGAAACGGCGCTTACCTGAGTCTGCTGGGCGTGCCTGATATGTGACGATTC	3420
Db	10970	GGCAGCGCCCGAGAAACGGCGCTTACCTGAGTCTGCTGGGCGTGCCTGATATGTGACGATTC	11022
OY	3421	CGCTTCGAGCCGGTTCGTTCGCGCAAGTGCGGCGCTGTGAGACGGGCGCGAGGGAGTGAACGTTC	3480
Db	11030	CGCTTCGAGCCGGTTCGTTCGCGCAAGTGCGGCGCTGTGAGACGGGCGCGAGGGAGTGAACGTTC	11088
OY	3481	GTGCTCAACTGCTCTCGGGGAGACTGATCGACAAAGATTTCATCTCTCGCATGCGCAC	3540
Db	11090	GTGCTCAACTGCTCTCGGGGAGACTGATCGACAAAGATTTCATCTCTCGCATGCGCAC	11149
OY	3541	GGCCGATTGTGAGCTTCGCGAAACGCGACTGTTTACGCGGATTAACAAGCTCGGCGTTCGG	3600
Db	11150	GGCCGATTGTGAGCTTCGCGAAACGCGACTGTTTACGCGGATTAACAAGCTCGGCGTTCGG	11208
OY	3601	CCGTTCCATCGGCATCTCTCCCTTCGCTGGTGGATCTCGGGGGATGATGCTTCGAGCGG	3660
Db	11210	CCGTTCCATCGGCATCTCTCCCTTCGCTGGTGGATCTCGGGGGATGATGCTTCGAGCGG	11268
OY	3661	CCGGCGCGGGGCTCGTGCCTCTTGGAGAGCTCTCCGCGCTTAATCGCGCAGCGCGTTC	3720
Db	11270	CCGGCGCGGGGCTCGTGCCTCTTGGAGAGCTCTCCGCGCTTAATCGCGCAGCGCGTTC	11322
OY	3721	ACCCCTCCCCCGAATCGGAGCGCTCCCGATGCGCCGTGTGCGCGATGCGTTCCGGAGCAGT	3780
Db	11330	ACCCCTCCCCCGAATCGGAGCGCTCCCGATGCGCCGTGTGTGCGCGATGCGTTCCGGAGCAGT	11388
OY	3781	GCGCAGGCGCAGCATCTTGGGAGACGTGTACTCACGCTGGGTGATCCCGAGAGTCCAGATC	3840
Db	11390	GCGCAGGCGCAGCATCTTGGGAGACGTGTACTCACGCTGGGTGATCCCGAGAGTCCAGATC	11448
OY	3841	CGTATTCCAAACCCACGAGGCGCGCGCCGCTCCACCGGGGATGGGACCTGTCTCGACAG	3900
Db	11450	CGTATTCCAAACCCACGAGGCGCGCGCCGCTCCACCGGGGATGGGACCTGTCTCGACAG	11508
OY	3901	CTCCGCTCAGCTGCGCGCGCGCGCGCGCGCGCTGTGAGGCGTTCCTCTGTAGCGAG	3960
Db	11510	CTCCGCTCAGCTGCGCGCGCGCGCGCGCGCGCTGTGAGGCGTTCCTCTGTGTAGCGAG	11568
OY	3961	GTCCTCGCAGGTGTCTGCGCAGCGCCCGAAATCAAGGTCCGGCGCGAGCGCTGTTACCCCG	4020
Db	11570	GTCCTCGCAGGTGTCTGCGCAGCGCCCGAAATCAAGGTCCGGCGCGAGCGCTGTTACCCCG	11622
OY	4021	CTCGGCATGAGCTGCGCTCATAGCCGTGGAGCTGTGGCAATGTATCGAGGCGAGCCTTCAG	4080
Db	11630	CTCGGCATGAGCTGCGCTCATAGCCGTGGAGCTGTGGCAATGTATCGAGGCGAGCCTTCAG	11688
OY	4081	CTGAAGCTGTGACGAGCGTTCTGTCTCAACGTCCCCCAATATGCGCTGTTGTGGCCCAAAAC	4140
Db	11690	CTGAAGCTGTGACGAGCGTTCTGTCTCAACGTCCCCCAATATGCGCTGTTGTGGCCCAAAAC	11749
OY	4141	CTGTTTGAATGCTCTTCGCGCACAGCTCTCTTGTGAGCGGGGTGGCGCGGAGAACTTACGG	4200
Db	11750	CTGTTTGAATGCTCTTCGCGCACAGCTCTCTTGTGAGCGGGGTGGCGCGGAGAACTTACGG	11808
OY	4201	GCAGGCGGCAAAACGACTTGTCTCATTCGAGCGGCGCACAATCAAGACTGGGAAATCATTTGCC	4260
Db	11810	GCAGGCGGCAAAACGACTTGTCTCATTCGAGCGGCGCACAATCAAGACTGGGAAATCATTTGCC	11866

QY	4261	CTATGA	4266
Db	11870	CTATGA	11875
<p>RESULT 6 US-09-568-472-1 ; Sequence 1, Application US/09568472 ; Patent No. 6358719 GENERAL INFORMATION: APPLICANT: Schnupp, Thomas APPLICANT: Lignon, James APPLICANT: Molnar, Istvan APPLICANT: Zirkle, Ross APPLICANT: Cyr, Devon APPLICANT: Goerlach, Joern TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOPHILONES FILE REFERENCE: 4-30852X CURRENT APPLICATION NUMBER: US/09/568,472 CURRENT FILING DATE: 2000-05-10 PRIOR APPLICATION NUMBER: 09/335,409 PRIOR FILING DATE: 1999-06-17 NUMBER OF SEQ ID NOS: 30 SOFTWARE: PatentIn Ver. 2.0 SEQ ID NO 1 LENGTH: 68750 TYPE: DNA ORGANISM: Sorangium cellulosum US-09-568-472-1</p>			
QY	Query, Match	100.0%;	Score 4266; DB 4; Length 68750;
	Best Local Similarity	100.0%;	Pred. No. 0;
	Matches 4266;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1	GTGGCGGATCTCCATCATGACGCGCGACCGAAGATCCGATTGGATCGTCGAGCGAGT	60
Db	7610	GTGGCGGATCTCCATCATGACGCGCGACCGAAGATCCGATTGGATCGTCGAGCGAGT	7655
QY	61	TGCGCTGTGCCCCGGTGGGTGATGATCTGAAGCGGTTCTGAGCGCTCTCGAGGGCTCG	120
Db	7670	TGCGCTGTGCCCCGGTGGGTGATGATCTGAAGCGGTTCTGAGCGCTCTCGAGGGCTCG	7729
QY	121	CGCGACACCGTCGGGCGGATCGCCCGCGGAAACGCTGGGATGACGAGCGTGGTTGATCC	180
Db	7730	CGCGACACCGTCGGGCGGATCGCCCGCGGAAACGCTGGGATGACGAGCGTGGTTGATCC	7785
QY	181	GACCCCGATGCCCCGGGGAGACGCGCCGTTACGCGCGCATCTTCTTGAGCGACGTAGCC	240
Db	7790	GACCCCGATGCCCCGGGGAGACGCGCCGTTACGCGCGCATCTTCTTGAGCGACGTAGCC	7848
QY	241	TGCTTCAGCGCTCTCTTGGGATCTGCGCTGCGGAAAGGCTGCGGATGGAACCTCGCA	300
Db	7850	TGCTTCAGCGCTCTCTTGGGATCTGCGCTGCGGAAAGGCTGCGGATGGAACCTCGCA	7908
QY	301	CATGACACTTGTCTGAGAGTGTGCTGGGAGCGCTGAGAAACGCGCGCATCGCTCATCG	360
Db	7910	CATGACACTTGTCTGAGAGTGTGCTGGGAGCGCTGAGAAACGCGCGCATCGCTCATCG	7965
QY	361	GCGCTCGTGGTACGGAACCGGAGTGTTCATCGGGATCGGCCCTTCGGAATATGAGGCC	420
Db	7970	GCGCTCGTGGTACGGAACCGGAGTGTTCATCGGGATCGGCCCTTCGGAATATGAGGCC	8029
QY	421	GCGCTCGCGGAAGGAGCGCGCTCGCGAGATCGACGCTCATGGCGGGCTGAGGAGCATG	480
Db	8030	GCGCTCGCGGAAGGAGCGCGCTCGCGAGATCGACGCTCATGGCGGGCTGAGGAGCATG	8088
QY	481	CCGACGCTCGGAGCGCGCGCATCTCGTATCCCTTCGGGCTGCGAGGGCCGTTGTTCGCG	540
Db	8090	CCGACGCTCGGAGCGCGCGCATCTCGTATCCCTTCGGGCTGCGAGGGCCGTTGTTCGCG	8148
QY	541	GTGGATACGGCCATATCTCTCTCGCTGGTGGCCCTTCATCTGCGCTGTCAAGCTTGGCG	600
Db	8150	GTGGATACGGCCATATCTCTCTCGCTGGTGGCCCTTCATCTGCGCTGTCAAGCTTGGCG	8206

QY 601 TCCGGGAATGCTCCACCGGCTGCTGGGATATCGTGAATTTGTCCGCGAGCAC 660
Db 8210 TCCGGGAATGCTCCACCGGCTGCTGGGATATCGTGAATTTGTCCGCGAGCAC 8269
QY 661 CTCTGTGGCTCTCGAAGACCCGGGCGCTGCGCAAGGACCGTCTGCAAGGCAATTTTCG 720
Db 8270 CTCTGTGGCTCTCGAAGACCCGGGCGCTGCGCAAGGACCGTCTGCAAGGCAATTTTCG 8329
QY 721 GCGGAGCGCGATGGGTTTGGAGCGAGCGAGGCGCTGCTGGTCTCTCAAGGCGCTC 780
Db 8330 GCGGAGCGCGAGGTTGGAGCGAGCGAGGCGCTGCTGGTCTCTCAAGGCGCTC 8389
QY 781 AGTGAAGCCCGCGCGAGCGCGATTCGATATTGGCGGTGATTGAGAGATCCGCGATCAAT 840
Db 8390 AGTGAAGCCCGCGCGAGCGCGATTCGATATTGGCGGTGATTGAGAGATCCGCGATCAAT 8449
QY 841 CACGACGGTGGAGCAGCGGTCTGACCGTGCCTGAAACGGAAGCTCCCAAGAAATCGTCTG 900
Db 8450 CACGACGGTGGAGCAGCGGTCTGACCGTGCCTGAAACGGAAGCTCCCAAGAAATCGTCTG 8509
QY 901 AAAGCGGCTGCTGGGAGCGAGCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 960
Db 8510 AAAGCGGCTGCTGGGAGCGAGCGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 8569
QY 961 GCGACCGGCGACGACCTTGTGATCCCAATCGAATCAAGCTCTGAAATGCGGTATACGAGC 1020
Db 8570 GCGACCGGCGACGACCTTGTGATCCCAATCGAATCAAGCTCTGAAATGCGGTATACGAGC 8629
QY 1021 CTCTGGGCGAGATGTTCGACACCGCTGCTGATCGGCTGCTGAGACCAACCTTGGCAT 1080
Db 8630 CTCTGGGCGAGATGTTCGACACCGCTGCTGATCGGCTGCTGAGACCAACCTTGGCAT 8689
QY 1081 CCTAGATATGGCTGGGAGATCACTGGGCTGCTGAGATGCTGCTGCTGCTGCTGCTGCTG 1140
Db 8690 CCTAGATATGGCTGGGAGATCACTGGGCTGCTGAGATGCTGCTGCTGCTGCTGCTGCTG 8749
QY 1141 CAGATTCCTGGGCACTTCACGCGCGAGCGCTGAACCCCGGATCTCAATGGGATGATCTT 1200
Db 8750 CAGATTCCTGGGCACTTCACGCGCGAGCGCTGAACCCCGGATCTCAATGGGATGATCTT 8809
QY 1201 CCGCTGACCTGTCACGCGCGCGCGACACCTGTGGCGCGACTGGAATACCGCGCGAGC 1260
Db 8810 CCGCTGACCTGTCACGCGCGCGCGACACCTGTGGCGCGACTGGAATACCGCGCGAGC 8869
QY 1261 GGGGAGCTGCTGCTGGGATGAGCGGAGCAACGCGGACGCTGGTCTGGAAGAGCGCGC 1320
Db 8870 GGGGAGCTGCTGCTGGGATGAGCGGAGCAACGCGGACGCTGGTCTGGAAGAGCGCGC 8929
QY 1321 GCGGCGACGTGTCACACCGCGCGCGCGAGCGACCGGAGAGCTGCTGGTCTGCGACA 1380
Db 8930 GCGGCGACGTGTCACACCGCGCGCGCGAGCGACCGGAGAGCTGCTGGTCTGCGACA 8989
QY 1381 AGGACCGGCTGACCGCTGATGACAGAGCGGCGGCTGCGGACCATCTGAGACCTTAC 1440
Db 8990 AGGACCGGCTGACCGCTGATGACAGAGCGGCGGCTGCGGACCATCTGAGACCTTAC 9049
QY 1441 CTTTTCGAGTGTCTGGGAGATGTGGCTTCACTCTGGGAGCGACCGCGAGCGGATGAG 1500
Db 9050 CTTTTCGAGTGTCTGGGAGATGTGGCTTCACTCTGGGAGCGACCGCGAGCGGATGAG 9109
QY 1501 CACCGGCTCGCGGTGGCGGACGTCGAGAGGAGGAGGCTGCGGAGCGCTGAGACGCTGCG 1560
Db 9110 CACCGGCTCGCGGTGGCGGACGTCGAGAGGAGGAGGCTGCGGAGCGCTGAGACGCTGCG 9169
QY 1561 GCGCAGGAGCAGACGTCGCGCGGTGGCGGTGCGGATCTGCGGATCTCTCAACGCGGAG 1620
Db 9170 GCGCAGGAGCAGACGTCGCGCGGTGGCGGTGCGGATCTGCGGATCTCTCAACGCGGAG 9229
QY 1621 CTCTGCTTCTTCTTCAACGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1680
Db 9230 CTCTGCTTCTTCTTCAACGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 9289

QY 1681 GATGATATGCTCCGCTTCCGAGAGCGTTGCACTGTGTCGTGAGGCTGTTCAACGAGAG 1740
Db 9290 GATGATATGCTCCGCTTCCGAGAGCGTTGCACTGTGTCGTGAGGCTGTTCAACGAGAG 9349
QY 1741 CTGCAACCGGCGCTTCCGAGAGGATGATGTGGGCGGAACCGGCGAGCGTGTGAGCGCGCTG 1800
Db 9350 CTGCAACCGGCGCTTCCGAGAGGATGATGTGGGCGGAACCGGCGAGCGTGTGAGCGCGCTG 9409
QY 1801 CTGCAACGAGCAGCTTCAACCGAGCGGCTGTTCACTTTCGAAATATGCGCTGCGCGC 1860
Db 9410 CTGCAACGAGCAGCTTCAACCGAGCGGCTGTTCACTTTCGAAATATGCGCTGCGCGC 9469
QY 1861 CTGTGCGGTGTGGGATGTAAGCCGAGATGTGTGCGCGGCAATAGCATCGTGAAGCTG 1920
Db 9470 CTGTGCGGTGTGGGATGTAAGCCGAGATGTGTGCGCGGCAATAGCATCGTGAAGCTG 9529
QY 1921 GTGGCTGCTGCGGTGGGAGGCTGTTCTGCTTGAAGACCGGAGTGTCTGTGGCTGCG 1980
Db 9530 GTGGCTGCTGCGGTGGGAGGCTGTTCTGCTTGAAGACCGGAGTGTCTGTGGCTGCG 9589
QY 1981 CCGGCGCGCTGATGACAGCGCTGCGGCGCGGCGGAGGAGTGTGATGAGGAGCGCG 2040
Db 9590 CCGGCGCGCTGATGACAGCGCTGCGGCGCGGCGGAGGAGTGTGATGAGGAGCGCG 9649
QY 2041 GAGGCGATGTGAGCTGTGCGGTGAGCGCGCGACGAGCGGTGCTGATGCGCGGCTC 2100
Db 9650 GAGGCGATGTGAGCTGTGCGGTGAGCGCGCGACGAGCGGTGCTGATGCGCGGCTC 9709
QY 2101 AACGCTCCGAGCAGAGTGTGATGCGGAGCGCGCGGCAACCGGTGATGCGAGTGGCG 2160
Db 9710 AACGCTCCGAGCAGAGTGTGATGCGGAGCGCGCGGCAACCGGTGATGCGAGTGGCG 9769
QY 2161 GCGATGGCGCGCGGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2220
Db 9770 GCGATGGCGCGCGGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 9829
QY 2221 CCGCTCATGAGCGCGAGTCTGAGAGCGGTTGCGGCGGTGAGCGAGTGTGAGTACCGG 2280
Db 9830 CCGCTCATGAGCGCGAGTCTGAGAGCGGTTGCGGCGGTGAGCGAGTGTGAGTACCGG 9889
QY 2281 CCGGCGGTGATGCTGCTGTGATGAGCAATTTGAGCGGAGGCTTGTGACAGAGAGTGGAGC 2340
Db 9890 CCGGCGGTGATGCTGCTGTGATGAGCAATTTGAGCGGAGGCTTGTGACAGAGAGTGGAGC 9949
QY 2341 TCGCGGAGCTATTTGAGGTCGCGACGCGGAGAGGAGTGTGCGCTTTCGAGATGAGTGAAG 2400
Db 9950 TCGCGGAGCTATTTGAGGTCGCGACGCGGAGAGGAGTGTGCGCTTTCGAGATGAGTGAAG 10009
QY 2401 GCGCTGCAACGCGGCGGCTGCGGAGCACTTGTGTGAGGTGCTGCGAATTCAGCGCTGCTC 2460
Db 10010 GCGCTGCAACGCGGCGGCTGCGGAGCACTTGTGTGAGGTGCTGCGAATTCAGCGCTGCTC 10069
QY 2461 GGCTGTGCTGCTGCTGCTGATGCGGAGCGCGCGGCGCTGCTGCTGCTGCTGCTGCTGCT 2520
Db 10070 GGCTGTGCTGCTGCTGCTGATGCGGAGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCT 10129
QY 2521 GGGGCTGACGAGCGGCGGAGCGGCTGCTGAGGCGCTGCGGCGGCTTGTGGGCGGTGGG 2580
Db 10130 GGGGCTGACGAGCGGCGGAGCGGCTGCTGAGGCGCTGCGGCGGCTTGTGGGCGGTGGG 10189
QY 2581 CTGTGCTTCTGAGGCGGCTTCTTCCCTCAGGAGGAGCGGCGGAGTGCCTGCTGCTGCT 2640
Db 10190 CTGTGCTTCTGAGGCGGCTTCTTCCCTCAGGAGGAGCGGCGGAGTGCCTGCTGCTGCT 10249
QY 2641 CTTTGGAGCGCGGAGCGCTACTGATGCAACGAAAGCTGAGAGCGGCGCGGCTGGGAGC 2700
Db 10250 CTTTGGAGCGCGGAGCGCTACTGATGCAACGAAAGCTGAGAGCGGCGCGGCTGGGAGC 10309
QY 2701 CCGGCTGCTGCGGAGCGGAGCAGAGCGAGGTGAGAGGAGGAGGCGGAGCGGAGCGGAG 2760
Db 10310 CCGGCTGCTGCGGAGCGGAGCAGAGCGAGGTGAGAGGAGGAGGCGGAGCGGAGCGGAG 10369
QY 2761 GACCGGCGAGCGGCTGCGCTGAGCAATTCGCGCGCGGAGCGGAGCGGAGCGGAGCGGAG 2820

Db 7790 GACCCGATGCCCCGGGAGACGCCCCGCTTACGCGCGCATCTTTCCTGAGGAGGTACC 7849
Qy 241 TGGTTGACGCGCTCTCTTCTTGCGCATCTCGCTCGAGAGCGCTGCGGATGAGCTTCA 300
Db 7850 TGGTTGACGCGCTCTCTTCTTGCGCATCTCGCTCGAGAGCGCTGCGGATGAGCTTCA 7909
Qy 301 CATGCACTCTTGCTGAGAGTGTGCTGAGAGCGCTGAGAGAGCGCGCGATGCTTCATCG 360
Db 7910 CATGCACTCTTGCTGAGAGTGTGCTGAGAGCGCGCTGAGAGAGCGCGCGATGCTTCATCG 7969
Qy 361 GCGCTCGTGGTACGAGAAAGGAGTCTTCACTCGGATCGCGCCGCTCGGAATATGAGGCC 420
Db 7970 GCGCTCGTGGTACGAGAAAGGAGTCTTCACTCGGATCGCGCCGCTCGGAATATGAGGCC 8029
Qy 421 GCGCTCGCGCAAGGAGCGCGCTCGGAGATCGACGCTCATGCGCGCGCTGAGGAGCAATG 480
Db 8030 GCGCTCGCGCAAGGAGCGCGCTCGGAGATCGACGCTCATGCGCGCGCTGAGGAGCAATG 8089
Qy 481 CCCAGCGTGGAGGCGCGCGCATCTGATGCTCGGCTCGGCTGAGAGCGCGCTGATGCGCG 540
Db 8090 CCCAGCGTGGAGGCGCGCGCATCTGATGCTCGGCTCGGCTGAGAGCGCGCTGATGCGCG 8149
Qy 541 GTGATACGCGCTTATCTCTCTCGCTGATGCGCTTCACTTGGCGCTGCAAGCTTGGCG 600
Db 8150 GTGATACGCGCTTATCTCTCTCGCTGATGCGCTTCACTTGGCGCTGCAAGCTTGGCG 8209
Qy 601 TCGCGGAGATGCTCAGAGCGCGCTGCGTGGTGGGATGCGTGAATGTTGTCGCGAGCAC 660
Db 8210 TCGCGGAGATGCTCAGAGCGCGCTGCGTGGTGGGATGCGTGAATGTTGTCGCGAGCAC 8269
Qy 661 CTCTGTGCTCTCTGAGAACCGGCGCGCTGCGCAAGGAGCGTCTGAGAGCAATTTTCG 720
Db 8270 CTCTGTGCTCTCTGAGAACCGGCGCGCTGCGCAAGGAGCGTCTGAGAGCAATTTTCG 8329
Qy 721 GCGGAGCGCGATGGCTTGAAGAGAGCGAGAGGCTGCTGCTGATGCTTCAAGCGCGCTC 780
Db 8330 GCGGAGCGCGATGGCTTGAAGAGAGCGAGAGGCTGCTGCTGATGCTTCAAGCGCGCTC 8389
Qy 781 AGTGAAGCGCGCGGAGCGCGGATCGGATTTGGCGGATTTGAGAGATCGCGGATCAT 840
Db 8390 AGTGAAGCGCGCGGAGCGCGGATCGGATTTGGCGGATTTGAGAGATCGCGGATCAT 8449
Qy 841 CACGACGCTGAGAGCGCGCTCTGACCTGCGCGAAGCGGAGCTCCAGAAATGCTGCTG 900
Db 8450 CACGACGCTGAGAGCGCGCTCTGACCTGCGCGAAGCGGAGCTCCAGAAATGCTGCTG 8509
Qy 901 AAACGCGCGCTGCGGAGCGAGGCTGCGCGCTTCTGCTGGGTTATGTCAGGCGACAC 960
Db 8510 AAACGCGCGCTGCGGAGCGAGGCTGCGCGCTTCTGCTGGGTTATGTCAGGCGACAC 8569
Qy 961 GCGACGCGGACGAGCGCTTGGTGAAGCCCATTCGAATCCAAAGCTTGAAGCGGTTATCGC 1020
Db 8570 GCGACGCGGACGAGCGCTTGGTGAAGCCCATTCGAATCCAAAGCTTGAAGCGGTTATCGC 8629
Qy 1021 CTGCGGCGAGATGTCGACGCGCGCTGCTGATTCGCGTGGTGGTGAAGACCACTTGGCCAT 1080
Db 8630 CTGCGGCGAGATGTCGACGCGCGCTGCTGATTCGCGTGGTGGTGAAGACCACTTGGCCAT 8689
Qy 1081 CTGAGTATGCTGCGGAGATCACTGAGGCTGCTGAAGGCTGCTTCTTCAAGCGAGG 1140
Db 8690 CTGAGTATGCTGCGGAGATCACTGAGGCTGCTGAAGGCTGCTTCTTCAAGCGAGG 8749
Qy 1141 CAGATTCTGCGGACCTTCAAGCGCGAGCGCTGAACCCCGGAGATTCATGAGGATTT 1200
Db 8750 CAGATTCTGCGGACCTTCAAGCGCGAGCGCTGAACCCCGGAGATTCATGAGGATTT 8809
Qy 1201 GCGGTCGACGCTCAGCGCGCGCGGACACGCTGCGCGGACTGGAATACCGCGCGAGCGG 1260
Db 8810 GCGGTCGACGCTCAGCGCGCGCGGACACGCTGCGCGGACTGGAATACCGCGCGAGCGG 8869
Qy 1261 GCGGTCGACGCTTCTGCGGATGAGCGGAGCAACGCGGACGCTGCTGGAAGAGCGCGG 1320
Db 1320 GCGGTCGACGCTTCTGCGGATGAGCGGAGCAACGCGGACGCTGCTGGAAGAGCGCGG 1320

Db 8870 GCGGTCGACGCTTCTGCGGATGAGCGGAGCAACGCGGACGCTGCTGGAAGAGCGCGG 8929
Qy 1321 GCGGTCGACGCTTCTGCGGATGAGCGGAGCAACGCGGAGCTGCTGCTGGAAGAGCGCGG 1380
Db 8930 GCGGTCGACGCTTCTGCGGATGAGCGGAGCAACGCGGAGCTGCTGCTGGAAGAGCGCGG 8989
Qy 1381 AGGACCGCGTCAAGCTTGAATGCAAGAGCGCGCGCTGCGGAGCAATCTGAGACCTTAC 1440
Db 8990 AGGACCGCGTCAAGCTTGAATGCAAGAGCGCGCGCTGCGGAGCAATCTGAGACCTTAC 9049
Qy 1441 CTTTCCGAGTGTCTGCGGAGATGAGCGCTTCAATGCTGCGGAGCAAGCGCGGATGAGAG 1500
Db 9050 CTTTCCGAGTGTCTGCGGAGATGAGCGCTTCAATGCTGCGGAGCAAGCGCGGATGAGAG 9109
Qy 1501 CACCGGCTCGCGGAGGAGCGAGCTGAGGAGAGGAGCTGCGGAGCGCTTGAAGCTTACG 1560
Db 9110 CACCGGCTCGCGGAGGAGCGAGCTGAGGAGAGGAGCTGCGGAGCGCTTGAAGCTTACG 9169
Qy 1561 GCGGAGGAGACAGAGTGTGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1620
Db 9170 GCGGAGGAGACAGAGTGTGCGCGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 9229
Qy 1621 CTGCGCTTCTCTTCAAGCGGAGAGGAGCGGAGCGCTGCGGAGCAATGAGCGCTTAC 1680
Db 9230 CTGCGCTTCTCTTCAAGCGGAGAGGAGCGGAGCGCTGCGGAGCAATGAGCGCTTAC 9289
Qy 1681 GATGATGATGCTGCGGAGGCTTCAAGCGGAGGCTTCAAGCGGAGGCTTCAAGCGGAGG 1740
Db 9290 GATGATGATGCTGCGGAGGCTTCAAGCGGAGGCTTCAAGCGGAGGCTTCAAGCGGAGG 9349
Qy 1741 CTGACCGGCGCTCTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1800
Db 9350 CTGACCGGCGCTCTGCGGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 9409
Qy 1801 CTGACCGAGACAGCTTCAAGCGGAGGAGGCTTCAAGCGGAGGCTTCAAGCGGAGG 1860
Db 9410 CTGACCGAGACAGCTTCAAGCGGAGGAGGCTTCAAGCGGAGGCTTCAAGCGGAGG 9469
Qy 1861 CTGAGCGGAGTGTGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
Db 9470 CTGAGCGGAGTGTGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9529
Qy 1921 GTGCGCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
Db 9530 GTGCGCTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9589
Qy 1981 GCGGCGCGCTGATGAGCGGCGCTGCGGCGGCGGCGGAGTGTGATGATGAGGCGCGG 2040
Db 9590 GCGGCGCGCTGATGAGCGGCGCTGCGGCGGCGGCGGAGTGTGATGATGAGGCGCGG 9649
Qy 2041 GAGGCGAGTGTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2100
Db 9650 GAGGCGAGTGTGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9709
Qy 2101 AACGCTCGGACAGAGTGTATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160
Db 9710 AACGCTCGGACAGAGTGTATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9769
Qy 2161 GCGATGCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Db 9770 GCGATGCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9829
Qy 2221 CCGGTCATGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
Db 9830 CCGGTCATGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9889
Qy 2281 CCGGTCGATGCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2340
Db 9890 CCGGTCGATGCTCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9949
Qy 2341 TCGCGGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2400
Db 9950 TCGCGGAGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10009

[illegible]

OY	3481	GTGCTCAACTCGCTCTCGGCGAGCTGATCGACAAGATTTCAAATCTCTGGATCGCAC	3540
Db	11090	GTGCTCAACTCGCTCTCGGCGAGCTGATCGACAAGATTTCAAATCTCTGGATCGCAC	11149
OY	3541	GGCCGCGTTTGTGAGAGCTCGGCAAGCCGACGTGTAACGCCGATTAACCAAGCTCGGCTGCGG	3600
Db	11150	GGCCGCGTTTGTGAGAGCTCGGCAAGCCGACGTGTAACGCCGATTAACCAAGCTCGGCTGCGG	11209
OY	3601	CCGTTCTGCGCAATCTCTCTTCTGCTGTGTGATCTCCGGGGATGATGCTTCGAGCCG	3660
Db	11210	CCGTTCTGCGCAATCTCTCTTCTGCTGTGTGATCTCCGGGGATGATGCTTCGAGCCG	11269
OY	3661	CCGGCGCGGGTCCGTGCGCTCTTGAGAGAGCTCTGAGCGCTGATCGCGGACAGCGTTC	3720
Db	11270	CCGGCGCGGGTCCGTGCGCTCTTGAGAGAGCTCTGAGCGCTGATCGCGGACAGCGTTC	11329
OY	3721	ACCCCTCCCCCATTCGAGCGCTCCGATCGCCGCTGTGCGCGATGCGTTCCGAGCATG	3780
Db	11330	ACCCCTCCCCCATTCGAGCGCTCCGATCGCCGCTGTGCGCGATGCGTTCCGAGCATG	11389
OY	3781	GCGCAGGCGCAGCATCTTGGGAAGCTGTACTGACGCTGAGGTGACCCGAGGTCGAGATC	3840
Db	11390	GCGCAGGCGCAGCATCTTGGGAAGCTGTACTGACGCTGAGGTGACCCGAGGTCGAGATC	11449
OY	3841	CGATTTCACACCAACGACGCGCGCGCGCGCGCTCCACCGGGGATCGGGACCTGCTGCACAGG	3900
Db	11450	CGATTTCACACCAACGACGCGCGCGCGCGCGCTCCACCGGGGATCGGGACCTGCTGCACAGG	11509
OY	3901	CTCGCGTCAGCTCGCGCGCGCGCGCGCGCGCTGAGGCGTTCTTCGTTACGCAAG	3960
Db	11510	CTCGCGTCAGCTCGCGCGCGCGCGCGCGCGCTGAGGCGTTCTTCGTTACGCAAG	11569
OY	3961	GTCCTGCGAGGTGCTGCGCACGCCGGAATCAAGGTGCGCGAGGCGCTTTCAACCGC	4020
Db	11570	GTCCTGCGAGGTGCTGCGCACGCCGGAATCAAGGTGCGCGAGGCGCTTTCAACCGC	11629
OY	4021	CTCGGATGGAATCTGCTCAATGAGCGGTGAGAGCTGCGGAATCGTATGAGGCGAGCGTCAAG	4080
Db	11630	CTCGGATGGAATCTGCTCAATGAGCGGTGAGAGCTGCGGAATCGTATGAGGCGAGCGTCAAG	11689
OY	4081	CTGAAGCTGTGACGAGAGTCTCTGTCAAGTCCGCCAATATGCGCTTGTGCGCCAAAC	4140
Db	11690	CTGAAGCTGTGACGAGAGTCTCTGTCAAGTCCGCCAATATGCGCTTGTGCGCCAAAC	11749
OY	4141	CTGTTGAGTCTCTCGGCACAGCTTCTCTTGGAGCGGAGTGGCGGAGAACTTACGG	4200
Db	11750	CTGTTGAGTCTCTCGGCACAGCTTCTCTTGGAGCGGAGTGGCGGAGAACTTACGG	11809
OY	4201	GCAAGCGTGCMAAACGATCTTGTCTATGCGGGCGAGATCAAGACTGGGAATCATTTGCC	4260
Db	11810	GCAAGCGTGCMAAACGATCTTGTCTATGCGGGCGAGATCAAGACTGGGAATCATTTGCC	11869
OY	4261	CTATGGA 4266	
Db	11870	CTATGGA 11875	
RESULT 8			
US-09-443-501A-2			
; Sequence 2, Application US/09443501A			
; Patent No. 6303542			
; GENERAL INFORMATION:			
; APPLICANT: Kosan Biosciences, Inc.			
; APPLICANT: Julien, Bryan			
; APPLICANT: Katz, Leonard			
; APPLICANT: Khosla, Chaitan			
; APPLICANT: Tang, Li			
; APPLICANT: Zierman, Rainer			
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing			
; FILE REFERENCE: 30062-20031.00			
; CURRENT APPLICATION NUMBER: US/09/443,501A			
; CURRENT FILING DATE: 1999-11-19			

PRIOR APPLICATION NUMBER: US 60/130,560
PRIOR FILING DATE: 1999-04-22
PRIOR APPLICATION NUMBER: US 60/122,620
PRIOR FILING DATE: 1999-03-03
PRIOR APPLICATION NUMBER: US 60/119,386
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: US 60/109,401
PRIOR FILING DATE: 1998-11-20
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 2
LENGTH: 71989
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 98.2%; Score 4187.6; DB 4; Length 71989;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 4217; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 GTGGGAGATCGTCCATCGAGCGCCGACGCCAAGATCCGATTCCGATGCTCGAGCGAGT 60
DB 1998 GTGGGAGATCGTCCATCGAGCGCCGACGCCAAGATCCGATTCCGATGCTCGAGCGAGC 2057
QY 61 TGCCTGTGCGCGGAGTGCATGATCGATCGAGCGGATTGAGACGCTCTCGAGGAGCTCG 120
DB 2058 TGCCTGTGCGCGGAGTGCATGATCGATCGAGCGGATTGAGACGCTCTCGAGGAGCTCG 2117
QY 121 CGCGACACCGTGCAGGAGTCCCGCCGACCGCTGGAGTGCAGACGCTGTTGATCC 180
DB 2118 CGCGACACCGTGCAGGAGTCCCGCCGACCGCTGGAGTGCAGACGCTGTTGATCC 2177
QY 181 GACCCGATGCCCCGGGAGAGCGCCCGTTAGCGCGCATCTTTCTTGAGCAGTACG 240
DB 2178 GACCTGATGCCCCGGGAGAGCGCCCGTTAGCGCGCATCTTTCTTGAGCAGTACG 2237
QY 241 TGCTTGAGCGCTCTCTTCTTGAGCATCTGCTCGGAGAGGCTCGAGTGAACCTGCA 300
DB 2238 TGCTTGAGCGCTCTCTTCTTGAGCATCTGCTCGGAGAGGCTCGAGTGAACCTGCA 2297
QY 301 CATCGACTTGTGCTGAGAGTGTGCTGAGAGCGCTGAGAGACGCCGATCGCTCCATCG 360
DB 2298 CATCGACTTGTGCTGAGAGTGTGCTGAGAGCGCTGAGAGACGCCGATCGCTCCATCG 2357
QY 361 GCGCTGTGCTGAGAGAGAGAGAGTGTGATCGGAGTGCAGCGCTCGCAATATGAGGCC 420
DB 2358 GCGCTGTGCTGAGAGAGAGAGAGTGTGATCGGAGTGCAGCGCTCGCAATATGAGGCC 2417
QY 421 GCGCTGCGGAGAGAGAGCGCGCTCGGAGATCGACGCTCATGAGGAGGCTGGGAGCGATG 480
DB 2418 GCGCTGCGGAGAGAGAGCGCGCTCGGAGATCGACGCTCATGAGGAGGCTGGGAGCGATG 2477
QY 481 CCCAGCTGAGAGCGGCGGAGATCTGTATGCCCTCGGAGCTGAGAGGCGCTGTCTCGG 540
DB 2478 CCCAGCTGAGAGCGGCGGAGATCTGTATGCCCTCGGAGCTGAGAGGCGCTGTCTCGG 2537
QY 541 GTGATGAGGAGCTTATCTGCTGCTGCTGAGGCGCTGATCATCTGAGCTGTGAGGCTGGC 600
DB 2538 GTGATGAGGAGCTTATCTGCTGCTGCTGAGGCGCTGATCATCTGAGCTGTGAGGCTGGC 2597
QY 601 TCCGGGAGATGCTCAACGCGCTGCTGCTGAGGAGTATGCTGATGCTGAGGAGCGACCC 660
DB 2598 TCCGGGAGATGCTCAACGCGCTGCTGCTGAGGAGTATGCTGATGCTGAGGAGCGACCC 2657
QY 661 CTGCTGTGAGCTCTCAAGAGACCGGAGCTGAGCGAGAGAGCTGCTGCAAGGAGCTTTG 720
DB 2658 CTGCTGTGAGCTCTCAAGAGACCGGAGCTGAGCGAGAGAGCTGCTGCAAGGAGCTTTG 2717
QY 721 GCGAGGCGGAGTGGTTCGAGAGAGGCGAGAGGAGTGGCGCTGTGAGTCTCAAGCGGCTC 780
DB 2718 GCGAGGCGGAGTGGTTCGAGAGAGGCGAGAGGAGTGGCGCTGTGAGTCTCAAGCGGCTC 2777

QY 781 AGTGAAGCCCGCGAGAGCGGATCGATATTTGGCGGTGATTGAGAGATCCGCGATCAAT 840
DB 2778 AGTGAAGCCCGCGAGAGCGGAGAGCGGATATTTGGCGGTGATTGAGAGATCCGCGATCAAT 2837
QY 841 CACGAGCGTGCAGAGCGGCTGTGACCGTGCCGAAAGGAGAGTCCCAAGAAATCGTCTG 900
DB 2838 CACGAGCGGAGAGAGCGGCTGTGACCGTGCCGAAAGGAGAGTCCCAAGAAATCGTCTG 2897
QY 901 AAAGCGGCGCTGGGAGAGAGAGGCGCGCGCTTGGGTGGGTATGAGGACACAC 960
DB 2898 AAAGCGGCGCTGGGAGAGAGAGGCGCGCGCTTGGGTGGGTATGAGGACACAC 2957
QY 961 GGCACGGGACGAGCGCTTGTGACCCCATCGAAATTCAGAGCTGTAATGCGGTATACGAC 1020
DB 2958 GGCACGGGACGAGCGCTTGTGACCCCATCGAAATTCAGAGCTGTAATGCGGTATACGAC 3017
QY 1021 CTGGGAGAGATGTCGACACCGCGCTGATTCGGGTGCTGAGAGACCAACTTGGCAT 1080
DB 3018 CTGGGAGAGATGTCGACACCGCGCTGATTCGGGTGCTGAGAGACCAACTTGGCAT 3077
QY 1081 CCGATGATGCGTGGGAGATCACTGGGCTGCGTGAAGGTGCTTGTCCCTTCAGACGGG 1140
DB 3078 CCGATGATGCGTGGGAGATCACTGGGCTGCGTGAAGGTGCTTGTCCCTTCAGACGGG 3137
QY 1141 CAGATTCTGCGCACTTCACGCGAGCGCTGAACCCCGGATCTCATGGGATGATCTT 1200
DB 3138 CAGATTCTGCGCACTTCACGCGAGCGCTGAACCCCGGATCTCATGGGATGATCTT 3197
QY 1201 CGGCTGACCTGACGCGCGCCCGGACACCTGTCGCGGACTGGAATACCCCGGACGAGG 1260
DB 3198 CGGCTGACCTGACGCGCGCCCGGACACCTGTCGCGGACTGGAATACCCCGGACGAGG 3257
QY 1261 GGGGTGACCTGTCGAGATGAGAGGAGGACCAACCGGACGCTGCTGAGAGAGGCGCG 1320
DB 3258 GGGGTGACCTGTCGAGATGAGAGGAGGACCAACCGGACGCTGCTGAGAGAGGCGCG 3317
QY 1321 GCGGCGAGTGCACACCGCGCGCGGAGGAGAGCGGAGAGCTGCTGCTGCGCA 1380
DB 3318 GCGGCGAGTGCACACCGCGCGCGGAGGAGAGCGGAGAGCTGCTGCTGCGCA 3377
QY 1381 AGGACCGGCTGACCGCTGAGATGACAGGCGGCGCGCTGCGGACCATCTGAGAGCTTAC 1440
DB 3378 AGGACCGGCTGACCGCTGAGATGACAGGCGGCGCGCTGCGGACCATCTGAGAGCTTAC 3437
QY 1441 CCTTCGAGTGTGAGGAGATGAGGCGCTTCACTGCTGAGAGAGCGGAGCGGATGAG 1500
DB 3438 CCTTCGAGTGTGAGGAGATGAGGCGCTTCACTGCTGAGAGAGCGGAGCGGATGAG 3497
QY 1501 CACCGCTCGGAGTGCAGGAGCTGAGAGAGGAGGCTGCGGAGAGCTTCTCAAGCGGCAAG 1560
DB 3498 CACCGCTCGGAGTGCAGGAGCTGAGAGAGGAGGCTGCGGAGAGCTTCTCAAGCGGCAAG 3557
QY 1561 GCGCAGGAGACAGCGTGCAGCGGCTGAGGAGAGTATCGCCGATCTCTCAAGCGGCAAG 1620
DB 3558 GCGCAGGAGACAGCGCGCGGCTGAGGAGAGTATCGCCGATCTCTCAAGCGGCAAG 3617
QY 1621 CTGCGCTTCTTCTTCAAGGAGAGAGGAGGAGAGAGTGTGAGGAGGCGGCTGTAC 1680
DB 3618 CTGCGCTTCTTCTTCAAGGAGAGAGGAGGAGAGAGTGTGAGGAGGCGGCTGTAC 3677
QY 1681 GATGATGATGCGGCTTCCGAGAGGCTTCAAGCTGAGGAGGCTGTTCAACAGAGAG 1740
DB 3678 GATGATGATGCGGCTTCCGAGAGGCTTCAAGCTGAGGAGGCTGTTCAACAGAGAG 3737
QY 1741 CTGACCGGCGCTTCGAGAGATGAGGAGCGAAGCGGACAGGCTGAGCGCGCGCTG 1800
DB 3738 CTGACCGGCGCTTCGAGAGATGAGGAGCGAAGCGGACAGGCTGAGCGCGCGCTG 3797
QY 1801 CTGACCGAGACGCTTCAACCAAGCGGCGCTGTTCACCTTCGAATATGCGCTGCGCG 1860
DB 3798 CTGACCGAGACGCTTTCACCAAGCGGCGCTGTTCACCTTCGAATATGCGCTGCGCG 3857

1861 CTGTGGCGGTCGTCGGGGTGTAGAGCGGAGTTGTTCGCGGCGCATGATCCGTGAGCTG 1920
DB |||||
3858 CTGTGGCGGTCGTCGGGGTGTAGAGCGGAGTTGTTCGTCGCGCATGATCCGTGAGCTG 3917
QY |||||
1921 GTGTGTCCTGTGGTCGCGGCGGCTGTCTCGCTTGAAGACGCGGTGTTCTCTGTGTCGCG 1980
DB |||||
3918 GTGTGTCCTGTGGTCGCGGCGGCTGTCTCGCTTGAAGACGCGGTGTTCTCTGTGTCGCG 3977
QY |||||
1981 CGCGGCGGCTGTGATGACGCGCTTCGCGCGCGCGGCGGCGATGTGTTCATCCAGCGCGCG 2040
DB |||||
3978 CGCGGCGGCTGTGATGACGCGCTTCGCGCGCGGCGGCGATGTGTTCATCCAGCGCGCG 4037
QY |||||
2041 GAGGCGGATGTGTGTCGTCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
DB |||||
4038 GAGGCGGATGTGTGTCGTCGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4097
QY |||||
2101 AACGCTCCGAGACGAGGTGTTCATCGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
DB |||||
4098 AACGCTCCGAGACGAGGTGTTCATCGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4157
QY |||||
2161 GCGATGCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2220
DB |||||
4158 GCGATGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4217
QY |||||
2221 CGGCTCATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280
DB |||||
4218 CGGCTCATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4277
QY |||||
2281 CGGCTCATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2340
DB |||||
4278 CGGCTCATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4337
QY |||||
2341 TGGCGGCGGCTTATGGGTGCGCGCACGCGCGGAGAGTGTGCGCTTCGCGGATGGAGTGAG 2400
DB |||||
4338 TGGCGGCGGCTTATGGGTGCGCGCACGCGCGGAGAGTGTGCGCTTCGCGGATGGAGTGAG 4397
QY |||||
2401 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2460
DB |||||
4398 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4457
QY |||||
2461 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2520
DB |||||
4458 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4517
QY |||||
2521 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2580
DB |||||
4518 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4577
QY |||||
2581 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2640
DB |||||
4578 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4637
QY |||||
2641 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2700
DB |||||
4638 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4697
QY |||||
2701 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2760
DB |||||
4698 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4757
QY |||||
2761 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2820
DB |||||
4758 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4817
QY |||||
4818 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4877
QY |||||
2881 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2940
DB |||||
4878 GGGCTGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4937
QY |||||
2941 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3000

DB |||||
4938 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4997
QY |||||
3001 CTGCCGGGAAACCCCACTTCCTGCTGTCTCGAGGCGAGTGTGCGCGGCGCATCTGTC 3060
DB |||||
4998 CTGCCGGGAAACCCCACTTCCTGCTGTCTCGAGGCGAGTGTGCGCGGCGCATCTGTC 5057
QY |||||
3061 GCGGTGGGCGAGGGGTGAACGCGCTCGTGTGTGGGCGCAACCGGTATAGCCCTTCGCGCG 3120
DB |||||
5058 GCGGTGGGCGAGGGGTGAACGCGCTCGTGTGTGGGCGCAACCGGTATAGCCCTTCGCGCG 5117
QY |||||
3121 GAGAGGTTGTCTAACCACTCAACAAGTGTGCGGTGTGTCGTCGTCGTCGTCGTCGTCGTCG 3180
DB |||||
5118 GAGAGGTTGTCTAACCACTCAACAAGTGTGCGGTGTGTCGTCGTCGTCGTCGTCGTCGTCG 5177
QY |||||
3181 CTCTGGGCGATGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3240
DB |||||
5178 CTCTGGGCGATGAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5237
QY |||||
3241 GACAGAAATAGCCCGCTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3300
DB |||||
5238 GACAGAAATAGCCCGCTTCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5297
QY |||||
3301 GTGCGTTCGCGCGGTCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3360
DB |||||
5298 GTGCGTTCGCGCGGTCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5357
QY |||||
3361 GGCAGCGCCGAGAAACGCGCTTACTGTGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3420
DB |||||
5358 GGCAGCGCCGAGAAACGCGCTTACTGTGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5417
QY |||||
3421 CGCTGCGACCGGTTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3480
DB |||||
5418 CGCTGCGACCGGTTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5477
QY |||||
3481 GTGCTCAACTGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3540
DB |||||
5478 GTGCTCAACTGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5537
QY |||||
3541 GGCAGCGGTTGTGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3600
DB |||||
5538 GGCAGCGGTTGTGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5597
QY |||||
3601 CGGTTCTGCGGAACT 3660
DB |||||
5598 CGGTTCTGCGGAACT 5657
QY |||||
3661 CGGCGGCGGAGTCCGTCTGTGAGGAGCTCTCGGCTGTGATCGCGGCGAGGCGGTCTTC 3720
DB |||||
5658 CGGCGGCGGAGTCCGTCTGTGAGGAGCTCTCGGCTGTGATCGCGGCGAGGCGGTCTTC 5717
QY |||||
3721 ACCCTTCCCCCATGTGCGACGCTCCGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3780
DB |||||
5718 ACCCTTCCCCCATGTGCGACGCTCCGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5777
QY |||||
3781 GCGGAGGCGGAGCATCTTGGGAGGCTCGATCAAGCTGGGAGTCCGAGGATCCAGATC 3840
DB |||||
5778 GCGGAGGCGGAGCATCTTGGGAGGCTCGATCAAGCTGGGAGTCCGAGGATCCAGATC 5837
QY |||||
3841 GGTATTCGAACCCAGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3900
DB |||||
5838 GGTATTCGAACCCAGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5897
QY |||||
3901 CTGCGGTACGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3960
DB |||||
5898 CTGCGGTACGTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5957
QY |||||
3961 GTCCTGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4020
DB |||||
5958 GTCCTGAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6017
QY |||||
4021 CTGCGCATGAGTCTGCTCATGAGCGGTCGAGTGTGAGGAGTGTGAGGAGGAGGAGGAGGAGG 4080

Db 6018 CTCGGCATGACCTGCTCATGCGCCGTGAGCTCGGCAATCGTATCGAGCGAGCTTCAAG 6077
Qy 4081 CTGAAGCTGTGACGACGCTTCTGTCAAGTCTCCCAATATCGCTTGTGGCCCAAAAC 4140
Db 6078 CTGAAGCTGTGACGACGCTTCTGTCAAGTCTCCCAATATCGCTTGTGGCCCAAAAC 6137
Qy 4141 CTGTTGATGCTCTGCGCACAGCTCTCTTGTGAGCGGGTGGCGGCGAGAACCTTAACG 4200
Db 6138 CTGTTGATGCTCTGCGCACAGCTCTCTTGTGAGCGGGTGGCGGCGAGAACCTTAACG 6197
Qy 4201 GCAGGCGCGCAAAACGACTTGTCTCATCGGGCGGAGATCAAGACTGGGAATATATGGC 4260
Db 6198 GCAGGCGCGCAAAACGACTTGTCTCATCGGGCGGAGATCAAGACTGGGAATATATGGC 6257
Qy 4261 CTATGA 4266
Db 6258 CTATGA 6263

RESULT 9
US-09-266-965-19
Sequence 19, Application US/09266965
Patent No. 6495348
GENERAL INFORMATION:
APPLICANT: Sherman, D
APPLICANT: Mao, Y
APPLICANT: Varoglu, M
APPLICANT: He, M
APPLICANT: Sheldon, P
TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
FILE REFERENCE: 600.456US1
CURRENT APPLICATION NUMBER: US/09/266.965
EARLIER FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: US 08/624,447
EARLIER FILING DATE: 1996-08-19
EARLIER APPLICATION NUMBER: PCT/US94/11279
EARLIER FILING DATE: 1994-10-06
EARLIER APPLICATION NUMBER: US 08/133,963
EARLIER FILING DATE: 1993-10-07
NUMBER OF SEQ ID NOS: 145
SOFTWARE: Parseq for Windows Version 3.0
SEQ ID NO 19
LENGTH: 3978
TYPE: DNA
ORGANISM: Streptomyces lavendulae
US-09-266-965-19

Query Match 19.1%; Score 814.2; DB 4; Length 3978;
Best Local Similarity 57.6%; Pred. No. 3.3e-134;
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

Qy 18 CGAGCCGCCACCGCAAGTCCGATTTGCGATGCTGAGAGCGAGTTGCCGTGCCCCGGTGG 77
Db 90 CGAGAGCGCGGCCCGCGAGCCCGTGCATGCTGGGAGTGGCGTGGCCCTTCCCGGGA 149
Qy 78 CGTATCGATCTGAGCGGGTCTGGAAGCTCTCGAGGGCTCGCGGACACGTCGAGGCG 137
Db 150 CGTGGCATCGCGGACGACCTGTGGCACTGTGTGCCAGGGCGGAGCGCGTCAACGA 209
Qy 138 AGTCCCGCCGAC---GCTGGATGACGAGCGGTGTTGATCCGACCCCGATGCCCC 194
Db 210 GTTCCCGCCGACCGGGGCTGGAGGTGAGCGCGCTTACAGACCCCGAGCGGACCCC 269
Qy 195 GGGGAGAGCGCCGCTTACGCGCGCATCTTCTGAGCGACGTAAGCTGCTTTCAGAGCTTC 254
Db 270 GGGGAGAGCGTACGCGCGCCACGCGGCGTCTTCAAGAGCGCGCGGATTTGAGCGCGC 329
Qy 255 CTTCTTGGCATCTGCTCGCTCGGAGCGCTGCGAGTGAACCTTGACATCGACTTTCCT 314
Db 330 CTTCTTGGCATCAAGCGCGCGGAGCGCTCGCCATGAGACCGAGCGCGCATGATCAT 389
Qy 315 GGAAGTGTGCTGGAGGCGCTGAGAGACCGCGCGATCGCTTCATCGGCGCTGTGGTAC 374

Db 390 GGAAGTCTCTGAGAGCGCTTTCAGACAGCGGCGCTTGACGCCACCCCTGCGGGCGA 449
Qy 375 GGAAGCGGAGTGTTCATCGGATCGGCGCGTCCGAATATAGCGCGCTGCGCAAGC 434
Db 450 GGAAGTGGGGTCTTCTGCTCGGCTTCAACAGAACGACTAAGTCAAGCTGTGAGCGC 509
Qy 435 GACGCGCTCGGACAGATTCAGCGCTCATGCGGGCTGGGAGCATGCCAGCTTGAGC 494
Db 510 GCGGAGTGGCGGAGGCGCTTCT-----ATCGGAGCGGCGCAATCCGCGAGCATCTCTC 563
Qy 495 GGGCGGAATCTGATATCCCTCGGCTGCGAGGGCGCTGCTGCGGATGATAGGGCTTA 554
Db 564 CGGCGGCTGCTTACACTTTCGCTTGAAGGCGCGGCGCTGCTGCTGAGACCGCTTG 623
Qy 555 TTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 614
Db 624 CTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 683
Qy 615 CACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 674
Db 684 CTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 743
Qy 675 GAAAGCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 734
Db 744 CCGCGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 803
Qy 735 GTTCCAGAGAGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 794
Db 804 CACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 863
Qy 795 GAGCGGATCGATGATTTGCGGATTCGAGATCCGCTGATCAATCAAGCGGTGCGAG 854
Db 864 CTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 923
Qy 855 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 914
Db 924 CGCGGCGCTGACCGCGCCCAACGAGCGGCGCAACGCGGTGATCCGCGAGGACCTGGC 983
Qy 915 GAGCGAGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 974
Db 984 CACGCAAGCTGACG 1043
Qy 975 GCTTGTGACCCCATCGAAATCCAGCTTGAATGCGGTATACGCGCTCGGCGAGATGT 1034
Db 1044 GCTGGGCGACCGCATTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1103
Qy 1035 CGCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1094
Db 1104 GGGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1163
Qy 1095 GGGGATCATGCGGCTGCTGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1154
Db 1164 CGGCGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1223
Qy 1155 CTTCCAGCGCGAGCGCTGTAACCCCGGATCTCAATGCGGTGATCTTCCGCTGACCGTCA 1214
Db 1224 GCTGCACTTCAAGAGCGCAACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1283
Qy 1215 GCGCGCGCGGACACCTGCGCGGAGTGAATACCGCGCGAGCGCGGCGGAGTGAAGCTGCT 1274
Db 1284 GACCGAGCGCGGAGATCGCGCGGACACCGGACACCGCGCGCGCGCGCGCGCGCGCGCTT 1343
Qy 1275 CGGATGAGCGGAGCAACGCGCGCGGCTGCTGGAAGAGCGCGCGCGCGCGCGCGCGCGCTG 1334
Db 1344 CGGCTGAGGAGCAACGCGCGCGGCTGCTGGAAGAGCGCGCGCGCGCGCGCGCGCGCGCTG 1403
Qy 1335 ACCGCG-----GCGCGGAGCGAGCGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1376
Db 1404 GAGCGCGCGCGCTGCGGAGCTGCTGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCTGCTG 1463
Qy 1377 GGCAGAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1436
Db 1464 CGGCGCGAGGAGCGCGCGCTGCTGAGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCTGCTG 1523

1437 CTACCTTGGAGTGTGTGGGCGATGTGGCTTCACTTGGGAGCGACCGCGAGCCAT 1496
1524 GCGCCCGACCTGGGACCGGCGGACCGGCTGGGACACCGCGGCGGCTT 1583
1497 GAGGACCGGCTGGGCTGGGCGGCGGAGTGTGAGGAGGCGGCTGGGCGGCTGGAGCG 1556
1584 GAGGACCGGCGGCGGCTGGGCGGCGGAGTGTGAGGAGGCGGCTGGGCGGCTGGAGCG 1643
1557 TGGGCGGCGGCGGAGCAGAGCTGGGCGGCGGCTGGGCGGAGTGTGGGCGGCTGGAGCG 1616
1644 CTGGGCGGCGGAGCAGAGCTGGGCGGCGGCTGGGCGGAGTGTGGGCGGCTGGAGCG 1703
1617 CAGGCTGGGCTTCTTCTTCAAGGAGGCGGAGGCGGAGGCTGGGCGGCTGGAGCG 1676
1704 CAGGCGGCGGCTTCTTCTTCAAGGAGGCGGAGGCGGAGGCTGGGCGGCTGGAGCG 1763
1677 GTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1736
1764 GCGGCTTCAAGGAGGCGGCGGCGGAGTGTGGGCGGAGGCTGGGCGGAGTGTGGGCGG 1823
1737 GAGGCTTCAAGGAGGCGGCGGCGGAGTGTGGGCGGAGGCTGGGCGGAGTGTGGGCGG 1796
1824 GAGGCTTCAAGGAGGCGGCGGCGGAGTGTGGGCGGAGGCTGGGCGGAGTGTGGGCGG 1883
1797 GGTGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 1856
1884 CTTGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 1943
1857 GCGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 1916
1944 GCGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 2003
1917 GGTGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 1976
2004 ACTGACCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2063
1977 TGGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2036
2064 GCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2123
2037 GCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2096
2124 GAGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2183
2097 GGTGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 2156
2184 GGTGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 2243
2157 GCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2216
2244 GCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2303
2217 CTACCGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 2276
2304 CTACCGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGG 2363
2277 CCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2336
2364 CCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2423
2337 GAGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGT 2396
2424 CCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2483
2397 GAGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2456
2484 CCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2543
2457 GGTGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGT 2513
2544 GAGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2603

2514 GCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2573
2604 GCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2663
2574 CCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2633
2664 CCGGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 2723
2634 CAGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGT 2680
2724 CAGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGT 2770

RESULT 10
US-09-266-965-74
; Sequence 74, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600.456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; EARLIER FILING DATE: 1993-10-07
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 74
; LENGTH: 12249
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-74

Query Match 19.1%; Score 814.2; DB 4; Length 12249;
Best Local Similarity 57.6%; Pred. No. 3.5e-134;
Matches 1549; Conservative 0; Mismatches 1100; Indels 30; Gaps 4;

18 CGAGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 77
1077 CGAGCGGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 1136
78 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 137
1137 CGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1196
138 AGTCCCGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 194
1197 GTTCCCGCGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 1256
195 GGGGAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 254
1257 GGGGAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 1316
255 CTTTCTTGGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 314
1317 CTTTCTTGGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGTGTGG 1376
315 GAGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGT 374
1377 GAGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGT 1436
375 GGAAGGAGGAGTGTTCATGAGGAGTGTTCATGAGGAGTGTTCATGAGGAGTGTTCATGAGG 434
1437 GAGGCTTCAAGGAGGCGGCTTCAAGGAGGCGGCGGAGTGTGGGCGGAGTGTGGGCGGAGT 1496

OY	435	GACGGCGCTCCGCAAGATCGACGCTCATGGGGGGCGTGGAGAACATGCCACGCTCGAGC	494
Db	1497	GGGGAGCGTCCCGCAGAGGGCTTC-----ATGGAGCCGGCAACTCCGCAAGATCTCTC	1556
OY	495	GGGGCGAATCTCGTATGCTTCGGGCTCCGAGAGGCGCTGTGCTCGCGTGGATACGACCTA	554
Db	1551	CGGCGCGCTGCCTCAACCTTTCGCTTGGAGAGGCCCGGCGCTGTCTCGTCAACCGCTTG	1610
OY	555	TTCTCTCTCGCTGTGGCCGTTTCATCTGGCTCTGACAGCTTGCGCTCCGGAGATGCTC	614
Db	1611	CTCTCTCTCGCTGTGGCTGTGGGCTATGCTGATTTGTCTGCGAGACCTCTGTGTGCTTC	1670
OY	615	CACGGGCTCTGTGTGTGGGCTATGCTGATTTGTCTGCGAGACCTCTGTGTGCTTC	674
Db	1671	CTGTGCGCTGTGGCGGGCGGGCGGACGCTATGAGCCAGCCGCTTTCATCGAGTTACG	1730
OY	675	GAAAGACCCGGGCGCTGTGGCCAGGAGACGTCGCTGCGACAAAGGCAATTTTGGGAGAGCCGATAG	734
Db	1731	CCGCGACGGGGGGCTGTGGCTCCCGAGAGCGCCCTGTGCAAGTCTTCTTGAGAGACCGCTGACG	1790
OY	735	GTTTGGAGCGAGGCGAAGGGTGCCTGCTGTGTCTCAAGCGGCTCACTGTGAGCCCGGC	794
Db	1791	CACCACTGTATCCGAGGGGCGGCGCGCTGTGCTGTGGCCCGGCTCTCGAGACGCCCGCG	1856
OY	795	GGAAGGCGCATGTGGATATTGGGGGTGATTTCAGAGATCCCGCATCAATCAACGCTGGAG	854
Db	1851	CCTGGGCTAACCCCGTGCACGCGGTCATCCGGGGCAGCCCGCTCAACAGAGACGGGCGAG	1910
OY	855	CAGCGGATGTGACCGTGGCCGAACGGAGAGCTCCCAAGAAATCGTGTGAAACGGGGCGTGGC	914
Db	1911	CGCGGGCTTGACCGGGCCCAACGACCGGCGCAACGCGGTATCCGGCAGGCACTGGC	1970
OY	915	GGAAGCGAGCTGCGCGCGCTTCTGGTGGATTATGTCCAGGCACACGCGACCGGACACAC	974
Db	1971	CAAGCGACGGGTGAAGCGCGCAACGGCTGACGCGGTCAAGGCAACACGCGACCGGACCCC	2030
OY	975	GCTTGTGTACCCCATCGAAATCCAACTCTGAATCGGATATACGGCTTCGGGCGAGATGT	1034
Db	2031	GCTGGGCGAACCCGATCGAGGCCCTGAGGCCCTCTCGCCACTTACGGGCGGGCCCGGCGGA	2090
OY	1035	CGCCACGCGCGTGCATCGGCTGGGTGGGTGAACCAACTTGGCCATCCTGAGTATGCTC	1094
Db	2091	GGGCAAGCGCGCTGTGGCTGGGCTGCTTAATCGAACTTGAGGCCACACCTCACTGCGCGC	2150
OY	1095	GGGATCACTAGGCTGTGAAGTGTCTTGTCTTCAGACAGGAGAGATTTCTTGCGCA	1154
Db	2151	CGGCGCGGGCGGCGTCAACAGATGTGATGCGCATGTGGGACAGGAGACCTGCCCGGAC	2210
OY	1155	CTTCACGCGGAGGCGCTGAACCCCGGATCTCATGGGGTATCTTCGGCTGACCGCTAC	1214
Db	2211	GCTGACCTCAAGGAGCCACCCCGCGGTGCATGTGTCGCCGGTGACGTACGGAGCTC	2270
OY	1215	GCGGCGCCCGGACACCGTGTGGCCCGGACTGGAATACGCTCGGACAGGGGTGGAGTGACTGTT	1274
Db	2271	GACCGAGGCGCCAGGACTGTGGCCGGAACCGGACAGCGCGGCGGTGGGCGCTCTGCTCTT	2330
OY	1275	CGGATGAGCGGAGCAACGCGGCACTGTGTGTAAGAGGCGCGCGGCGGACAGTGTGAC	1334
Db	2331	CGGCTCAAGCGGACCAACGCTCATGTATCTGTGAGGGGCCCGCCCGGAGAGGACAC	2390
OY	1335	ACCGCG-----GGCGGAGAGCACCGGCAAGCTGTGTGTGTCTC	1376
Db	2391	GAGCGCGCGCTGGCGGACGTCTCTCGACGCGCGGGGCCCGCTGCGGTGGTGTCTC	2450
OY	1377	GGCAAGACCGCGCTCAAGCTCTGTGAATGCAAGGCGCGCGGTGGCGGACCATGTGGAGAC	1436
Db	2451	CGGCGCGAGCGGCGCGCTGTCCGAGCGCAAGCCAGGCGCTGTGGGCGCCACTGTGACGC	2510
OY	1437	CTACCTTTCGAGTGTCTGTGGCGATGTAGGCTTCACTGTGCGAGACGACCGCAAGCGCAT	1496
Db	2511	GCGCCCGCACTGTGACACCGGCGGACGTGGCCACCGCGCTGTGCGACACACCGGGCGGCTTT	2570
OY	1497	GGAGCACGGCTCGCGGTGGCGGCGACGTGAGGAGGGGCTGTGCGGACGCGCTTGTGACG	1556

Db	2571	CGACCACTGGGGCCGCGCTCGTCGGCCGGGAACCTGAGGAATTGCTCGCGGACTTCGCGGC	2630
Qy	1557	TGCGCGCAGGAGACAGACGTCGCCCGGTGCGGTGCGGAGTATTCGCGCAATTCCTACGCGG	1616
Db	2631	CTTGCCACGGGAAACCGCGCGGCCCGCGCTGGTCAACGCGGCGGAAACCCGCGCTCGCGGG	2690
Qy	1617	CAAGCTGCGCTTTCTTTACCGGACAGAGGGGCGACAGCTGAGCATGGCGCTGGAGCT	1676
Db	2691	CAAGGCGCCCTCTCTTCAACGGAACAGGGCAGCAGACGACCCCGGCAATGGCGCGCAACT	2750
Qy	1677	GTAGAGTATATGTCCTCGGTTCCCGGAGGCGTTGACCTGTGCGTAGAGCTGTTCAACA	1736
Db	2751	GCGGCTCACAGCACCGGTTCGCCGAGCGCTGAGACAGGTATGCGCCACGCTGACCG	2810
Qy	1737	GGAGCTCGACCGGCGGCTCCGCGAGGAGTATGAGGCGGAACCGGCGACGCTGACGCGCG	1796
Db	2811	GCACCTCGAACCGGCGCGCTGCGCGAGGTCTGTTGCGCGGAGACGGACGCCGAGCGCG	2870
Qy	1797	GCTGCTGACACAGACAGCTTACCCACGCGCGGCTGTTACCTTGGAAATGCGCTCGC	1856
Db	2871	CTGCTCTGACACGACGCGCTACCCAGCGCGGCTGTTCCGCTCGAGGTGCGGCTGCT	2930
Qy	1857	CGCGCTGTGGGGTCTGAGGGGTATGAGCGCGAGTGTGCGCGGCAATACACATCGTGA	1916
Db	2931	GCGGCTGTGAGGACATGGGGCTTGGCGGCGCGGATGATCGCGGACATTCGCTGGGA	2990
Qy	1917	GCTGATGAGCTCTGCGGTGGGCGGCGGTTCCTGCTTGAAGAACCGGCTTCTGATGAGC	1976
Db	2991	ACTGACCGCGCGCTACCGCGCGCGGAGTCTGCTGCTGCGGACGCTGCGCGCTGCTGCG	3056
Qy	1977	TGCGCGCGGCGGCTGATGACAGGCGCTGCCGCGCGGCGGCGAGTGTGATGACAGGC	2036
Db	3051	CGCCCGCGGCGGCTGACCCAGGACATGCCCGCGGCGGCGCATGAGTCCGCTGACAGGC	3110
Qy	2037	GCGGAGGCGGATGTGGCTGTGCGGTGGCGGCGGACGACGCTGCGTGTGATTCGCGC	2096
Db	3111	GACCGAGGACAGGATGGCGCGGCCCACTGCCGAGCGCGCGCGCGGAGCATTCGCGCG	3170
Qy	2097	GSTCAACGCTCCGAGCCAGGAGTGCATCGCGGCGGCGGCGAACCCGCGATGAGATGCG	2156
Db	3171	CGTACCGGACCGGAAGGATGTGTGCTCTCGGCGACGAGCGCGCTGACGACCTGGC	3230
Qy	2157	GCGGCGATGAGCGCGCGCGGCGGCGGACCAAGCGGCTCCACGCTTCGATGCGATTCCA	2216
Db	3231	GCGGAGTGGGCGCGCGCGCGGAGAACAGAGGCTGGGGTCAAGCACGCGCTTCCA	3290
Qy	2217	CTCACCGCTCATGCGCCCGATGCTGAGAGCGTTGCGGCGTGTGCGCGGATCGTGAAGTA	2276
Db	3291	CTCGGCCACCTGGAACGCAATGACCGAGGCGTTGCGCGAGGTGCCACGAGGGGTCTCTA	3350
Qy	2277	CGGCGCGCGGTGCATGCTGCTGTGCACGAATCTGAGGGGAAAGGCTTGCACAGACGAGT	2336
Db	3351	CAGGCGCGCTCTCCCGGTGTCTCAGCTACCGGGGCCCCCGTCACTGACGAGAGCT	3410
Qy	2337	GAGCTCGCGCGGCTATTGGGTGCGCCACGCGCGAGAGGTGTGCGCTTCCGCGATGAGT	2396
Db	3411	CCGCAAGGCGGAAACATGGGTGCGGCGACGTCCGGGAAAGCGTGTGCTTCCACGACGGGT	3470
Qy	2397	GAAGCGCTGACGAGCGGCGGCTGCGGACACTTGTGTGAGGTGCGTCCGAATGACGCT	2456
Db	3471	CGGCGCTCTGCGCACGCGCGGCGCACCGCGCTTCTGAGAGGTGCGGCGCGCGCGTCT	3530
Qy	2457	GCTGGGCGCTGTGCTGCTGCTGACATGCGGAGACGCCGCGC---GGCGGCTCGGATGTC	2513
Db	3531	GACGCGCGCGGACCGGAGTCTTGCGCGAGCGCGCCCGAGAGCTTGTCTCCGCTGCT	3590
Qy	2514	GCGCGCTGAGCGTACAGCGCGGCGGACCGTGTCTGAGAGCGCTTCGCGGCGCTTGGGCGT	2573
Db	3591	GCGGCGCGGACCGGACCCAGATCGTGTCTGACGCGCGTGGCGGACGCCACAGAT	3650
Qy	2574	CGGTGGCTGTCTCTGGGCGGCGCTTTCCTTCAGGAGGAGCGCGGAGTGGCGCTGCC	2633

Db 3651 CGGCGCTGCGCGGGGTGGAGCGGCTGCTGCGCAAGCGCCGGAGCGCGGTGACCTGCG 3710
QY 2634 CACGTAACCTTGGCAGCGCGGAGCGCTACTGATTCGACACAGAAAGCCG 2680
Db 3711 CACGTAACGCTTCCAGCGCGGCGGACTACTGCTGGCGGCGAGTGGCG 3757
RESULT 11
US-09-266-965-96
; Sequence 96, Application US/09266965
; Patent No. 6495348
; GENERAL INFORMATION:
; APPLICANT: Sherman, D
; APPLICANT: Mao, Y
; APPLICANT: Varoglu, M
; APPLICANT: He, M
; APPLICANT: Sheldon, P
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster
; FILE REFERENCE: 600,456US1
; CURRENT APPLICATION NUMBER: US/09/266,965
; EARLIER FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 08/624,447
; EARLIER FILING DATE: 1996-08-19
; EARLIER APPLICATION NUMBER: PCT/US94/11279
; EARLIER FILING DATE: 1994-10-06
; EARLIER APPLICATION NUMBER: US 08/133,963
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 96
; LENGTH: 18311
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
; US-09-266-965-96

Query Match 19.1%; Score 814.2; DB 4; Length 18311;
Best Local Similarity 57.6%; Pred. No. 3,6e-134;
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

QY 18 CGAGCGCGGAGCGGAGATCCGATTCGATCGTCGAGCGGAGTGGCGTCCGTCGCGGTGG 77
Db 1077 CGAGACGCGGCGCGCGGAGCGCGGCGGATCGTCGATCGTCGCGGATCGTCGCGGCGG 1136
QY 78 CGTGAATCATCTGACGCGGCTTCTGACGCTCTCCAGAGGCTGCGGAGAACGCTGGGGG 137
Db 1137 CGTGCATCTGCGAGCGACTGTGTGCGAGCTGTGCGCGAGGCGCGGAGCGCGTCAACGA 1196
QY 138 AGTCCCGCGGAAAC--GCTGGAGATGACAGCGTGTGTTGATCCCGAACCCGATGCC 194
Db 1197 GTTCCCGCGGAGCGGCGGCTGGAGCGTCAACGCTGTACGACCCGAGACCGGGGACCC 1256
QY 195 GGGGAAAGCGCGCTTACGCGCGCATCTTCTGAGCGAGTACGCTGCTTGGACGCTC 254
Db 1257 GGGGAGAGCGTACGCGCGCACGCGCGCTTCTCAAGAGCGCGCGGATTCGAGCGCGC 1316
QY 255 CTTCTTGGCATCTGCGCTCGGAGCGCTGCGAGTGGACCTTGACATCGACTCTGT 314
Db 1317 CTTCTTGGCATCTGCGCGCGCGGAGCGCTGCGCATGAGCGCGAGCGCATGATCAT 1376
QY 315 GAGAGTGTGCTGGAGAGCGCTGAGAAAGCGCGCATGCTCCATCGGCGCTGCTGCTAC 374
Db 1377 GAGAGTGTGCTGGAGAGCGCTTGAAGAGCGGCGCTTCAAGCAACCTTGGGGGCGA 1436
QY 375 GGAAGCGGAGTGTTCATCGGAGTGGCGCGCTTCCGATATGAGCGCGCGTCCGCAAGC 434
Db 1437 GGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1496
QY 435 GACGCGCTGCGGAGATGAGCGCTCATGAGCGGCGCTGGGAGAGTCCAGAGTGGAGC 494
Db 1497 GCGGAGCGTGGCGGAGGCTTC-----ATCGGAGCGGCACTCGCGAGATCTCTTC 1550
QY 495 GGGCGCAATCTGTATGCGCTGCGGCTGCGAGGCGCGTGTGTGCGAGTATACGAGCTTA 554

Db 1551 CGGCGCGGTGCGCTACACCTTCCGCTTCCAGAGGCGCGCGCGCTGCTGACACCGCGCTG 1610
QY 555 TTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 614
Db 1611 CTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1670
QY 615 CACGCGCGTGGCTGGGTATGCGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674
Db 1671 CTTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1730
QY 675 GAAAGCGCGGCGCTGCGGAGAGAGTGTGCTGCAAGGCACTTTTGGCGGAGCGGATG 734
Db 1731 CCGCGAGGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1790
QY 735 GTTGGAGAGAGCGGAGGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 794
Db 1791 CACCACTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1850
QY 795 GAGCGCGATCGGATATTTGGCGGATTCGAGGATTCGCGATCAATCAAGCGGCGGCGG 854
Db 1851 CTTGGGCTACCGCTGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1910
QY 855 CAGCGCTGACCGCTGCGGAGAGTCCCAAGAAATGCTGTAAGAAAGCGGCGCTGCG 914
Db 1911 CCGCGGCGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1970
QY 915 GAGCGGAGCTGCGCGCGCTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974
Db 1971 CAACGCGAGGCTGAGCGGCGGAGCGGCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2030
QY 975 GCTTGTGATCCCATCGAATTCGAAATTCGAAATTCGAAATTCGAAATTCGAAATTCGAAATTCG 1034
Db 2031 GTTGGGCGAATCCGATTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2090
QY 1035 CGCGAGCGCGCTGCTGATCGGCTGCTGAGAGCAACCTTGGGCACTCTGATGCTGCTG 1094
Db 2091 GGGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2150
QY 1095 GGGGATCACTGCGCTGCTGAGAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1154
Db 2151 CGGCGCGGCGGCGCTGATCAAGATGTGTATGCGGATGCGGAGCGGAGCGGCGGCGGCGGCGG 2210
QY 1155 CTTCCAGCGGAGGCGTGAACCGCGGATTCATGTGAGGTATCTTGGCTGACCGCTAC 1214
Db 2211 GTTGAATCTCAAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2270
QY 1215 GCGCGCGGAGCAACGCTGCGGAGTGAATACCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 1274
Db 2271 GACCGAGGCGGAGAGCTGCGGAGCAACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2330
QY 1275 CGGATGAGCGGAGCAACCGGAGAGTGTGCTGAGAGAGCGCGGCGGCGGCGGCGGCGGCGG 1334
Db 2331 CGGCTGAGCGGAGCAACCGGAGAGTGTGCTGAGAGAGCGCGGCGGCGGCGGAGGAGCAAC 2390
QY 1335 ACCGCG-----GCGCGGAGCGGAGCGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1376
Db 2391 GAGCGCGCGCTGCGGAGAGCTCTCTGCGAGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 2450
QY 1377 GGCAGAGACCGCGTACCGCTGATGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1436
Db 2451 CGGCGGAGCGGAGGCGGCTGCGGAGCGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2510
QY 1437 CTACCTTTCGAGTGTCTGCGGAGTGTGCGGCTTCACTTGTGCGAGAGCGGCGGCGGCGG 1496
Db 2511 GCGCGCGGAGCTGCGGAGCGGAGTGTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2570
QY 1497 GAGAGCGGCGCTGCGGAGTGTGCGGAGCGTTCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1556
Db 2571 GACCAACCGGCGGCGGCTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2630
QY 1557 TCGCGCGAGGAGAGAGAGTGTGCGGCGGAGTGTGCGGAGTGTGCGGAGTGTGCGGAGTGTGCG 1616
Db 2631 CTTGCGCAACCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2690

1617 CAAGCTCCCTTTCTTCTTACCGGACAGGGGCGCAGACGCTGCGCATGCGGCGGTGGCT 1676
Db CAAGGCGCGCTTCTTCTTACCGGACAGGGGCGCAGACGCGGCGCATGCGGCGGTGGCT 2750
Qy 1677 GTACGATGTTATGTTGCTTCCGCGGAGCGCTTCACTGCTGCTGAGGCTGTTCAACA 1736
Db 2751 GCGCGCTCACAGCACCGGTGTTCCGCGAGCGCTTGACAGGCTTCCGCGGCGGTGAGCG 2810
Qy 1737 GGAGCTGACCGGCGGCTCCGCGAGGATGTTGGCGCAACCGCGCAGGCTGACGCGCG 1796
Db 2811 GCACCTGACCGGCGGCTCCGCGAGGATGTTGTTCCCGCGAGCGCAGCGCGCGCGCG 2870
Qy 1797 GCTGCTGACGACAGACGCTTCAACCGCGCGGCTGTTCACTTCAATATGCGCTGCG 1856
Db 2871 CTTGCTGACAGACGCGCTTCAACCGCGCGGCTGTTCCCGCGAGGCTGCGCGCTGCT 2930
Qy 1857 GCGGCTGTTGCGGCTGTTGAGGCTGTTGAGCGGAGTTGTTGCGCGCGCATGAGTGTGA 1916
Db 2931 GCGGCTGTTGAGGACTGCGGCTTCCGCGCATGTTGCGCGCGCATGCTGCGCGCA 2990
Qy 1917 GCTGCTGCTGCTGCTGCGGCGGCGGCTTCTCCGCTTGAAGACGCGGCTTCTGCTGCG 1976
Db 2991 ACTGACCGCGCTTACGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3050
Qy 1977 TGCGCGCGGCGCGCTGATGACAGCGCGCTGCGCGCGCGCGCGATGTTGATGAGCG 2036
Db 3051 GCGCGCGCGCGCGCTGATGACAGCGCGCTGCGCGCGCGCGCGATGTTGATGAGCG 3110
Qy 2037 GCGCGAGCGCGATGTTGCTGCTGCGGCGCGCGCGCGCGCGCGCGATGTTGATGAGCG 2096
Db 3111 GACCGAGGACGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3170
Qy 2097 GGTAAACGCTCCGACGAGGCTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2156
Db 3171 GGTAAACGCTCCGACGAGGCTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3230
Qy 2157 GCGCGCGATGCG 2216
Db 3231 GCGCGAGGCG 3290
Qy 2217 CTACACGCTCATGCG 2276
Db 3291 CTACACGCTCATGCG 3350
Qy 2277 CCG 2336
Db 3351 CAGCG 3410
Qy 2337 GAGCTCG 2396
Db 3411 CCGCGAGCG 3470
Qy 2397 GAAAGCGCTGACG 2456
Db 3471 CCG 3530
Qy 2457 GCTGCG 2513
Db 3531 GACG 3590
Qy 2514 GCG 2573
Db 3591 GCG 3650
Qy 2574 GCGTGGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2633
Db 3651 GCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3710
Qy 2634 CACGTAACCTTGGAGCG 2680
Db 3711 CACGTAACCTTGGAGCG 3757

RESULT 12
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Beljach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziemann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144.085
; EARLIER FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010.809
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 33529
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-144-085-3
Query Match 18.8%; Score 803.2; DB 3; Length 33529;
Best Local Similarity 57.6%; Pred. No. 3.1e-112;
Matches 1591; Conservative 0; Mismatches 1128; Indels 41; Gaps 7;
Qy 37 CCGATTGCGATGTCGCGAGCGAGTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 96
Db 25316 CCGATTGCGATGTCGCGAGCGAGTTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 25375
Qy 97 TTCTGAGCGCTCTGCGAGCGCTCCGCGGACACCGTCCGCGCGCGCGCGCGCGCGCG 156
Db 25376 TTCTGAGCGCTCTGCGAGCGCTCCGCGGATACGCTGCGCGCGCGCGCGCGCGCG 25435
Qy 157 GATGACGAGCGTGTGTTGATCCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 216
Db 25436 GATGACGAGCGTGTGTTGATCCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25495
Qy 217 GCATCTTCTGACGCGAGTACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 276
Db 25496 GCATCTTCTGACGCGAGTACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25555
Qy 277 GAAAGCGCTGCGAGCG 336
Db 25556 GAAAGCGCTGCGAGCG 25615
Qy 337 GAGAACGCGCGATCGCTCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 396
Db 25616 GAGAACGCGCGATCGCTCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25675
Qy 397 ATCCG 456
Db 25676 ATCCG 25726
Qy 457 GCTCATGCG 516
Db 25727 GCTCATGCG 25786
Qy 517 GGGCGTGGAGGCGCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 576
Db 25787 GGGCGTGGAGGCGCGTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25846
Qy 577 CATCTGCGCTGTCAGAGCTTGGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 636
Db 25847 CATCTGCGCTGTCAGAGCTTGGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 25906
Qy 637 TCGCTGATGTTGTCGCGAGCACCTGTGTGCTGCTGCGAGACCGCGCGCGCGCGCG 696
Db 25907 TCGCTGATGTTGTCGCGAGCACCTGTGTGCTGCTGCGAGACCGCGCGCGCGCGCG 25966

QY 697 GACGCTGCTGCAAGGCAATTTTCGCGGAGGCGCATGSGTTCCGACGAGCCGAAAGGTC 756
DB 25967 GACGGGCGGTGCAAGGCGTTCTCGACGGGCGGAGGTTATGCGCTGCGGAGGAGTC 26026
QY 757 GCGCTGCTGCTCTCAAGGCGCTCAGTGGAGCCCGCGGAGCGGCGATGGATATATGGC 816
DB 26027 GGGGTCTGCTGCTGATGCGGCTGTCGAGGCGGAGCGGAGGAGGCGGCTGCTGCTG 26086
QY 817 GTGATTCGAGATCCGCGATCAATCAACGAGTGGCGAGCGGCTGACCGTCCGGAAC 876
DB 26087 GTGGTGGCGGACGCGCGGCTCATGAGACGGGCGGAGCGGAGTCAAGGCGCGGAC 26146
QY 877 GGGAGCTCCCAAGAAATGCTGCTGAACGGGCGCTGAGCGAGCGGAGGCTGCGCGCT 936
DB 26147 GGCACGGCGCAGAGAGGTGTGCGCGGCGCTGCGAGACGCGGCGCTGAGCGCGCG 26206
QY 937 TCGGTGGGTATATGTCGAGGACACGCGGACGAGGCGCTTGTGTGACCCCATCGAATC 996
DB 26207 AGCATCGATGTGTGAGTGCACGCTACGGGCACTGCTGGCGACCCGATCGAGTG 26286
QY 997 CAAGCTTGAATGCGGTATACGACCTCGGCGAGATGTGCGACGCGCTGCTGATCGG 1056
DB 26267 CAGGCGCTCGGCGGCTGACGGGCAAGGTGCGGATATGCTGCTGCTGCGAGCTGGC 26326
QY 1057 TGGGTGAAGAACCACTTGGCCATCTGAGTATGCTGCGGAGATCATGGGCTGCTGAG 1116
DB 26327 GCGGTCAAGAGCAATATCGGTATCTGAGTCCGCGCGGAGTCCAGAGGAGTGTCAAG 26386
QY 1117 GTGCTTGTCTCTTCAGACGCGGCGAGATTCCTGCGCATCTCAAGCGGCGCTGAGAC 1176
DB 26387 ATCTGTGGCGGCTTCTGTTACGATGCTGCGCGGACGCTGACGCTGCGCGCAAT 26446
QY 1177 CCGCGATCTCATGGGTGATCTTGGCTGACCGCTGACGCGCGCGGACACCGTGGC- 1235
DB 26447 CCGCGATCTCTGGGAGAGACTGCGCGGTGCAGTGCTGCTGACCGCTGCGCT 26506
QY 1236 -----GACGTGAATACCGCGGAGCGGCGGAGTGAAGTCTGTCGAGTGAAGCGGAGC 1290
DB 26507 CCGCGCGAGAGAGGCGCGCGCGGTGCGGTGCTGCTGCTGAGATCAGCGGAGC 26566
QY 1291 AACCGGCACTGCTGAGAGAGGCGCGGCGGCGACCTGACACCGC-----CG 1342
DB 26567 AACCGGCACTGCTGAGAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 26626
QY 1343 CCGCGGAGGACCGGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1402
DB 26627 AGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 26686
QY 1403 CACACGCGGCGGCGGCTGCGCGGACCATCTGAGAGCACTACCTTTCGAGTGTGCGGAGT 1462
DB 26687 CCGAGGCGGCGGCGGCGGAGTGTGAGAGCACTGAGAGCACTGAGAGTGTGCGGAGT 26746
QY 1463 TGGCTTTCAGTCTGAGGACGAGCGGCGGCGGATGAGGACCGGCTGCGGCTGCGGCGA 1522
DB 26747 TGGTCCGACGCGGCGGCTGCGACCGGACCGACCTTGAAGTGGGCGGCTGCTGCTGCGG 26806
QY 1523 CGTCAAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1582
DB 26807 CGAGGCGCTGCGGAGCGTGTGAGGCGGCTTTCGCGCGCTGCTGCGGCGGCGGCGGAG 26866
QY 1583 GTGCGGTGCGAGTATCGCGGATTCCTCAAGGCGGAGGCTGCGCTTCTTCAAGGAGC 1642
DB 26867 CGGTGTGAGCGGAGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 26923
QY 1643 AGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1702
DB 26924 AGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 26983
QY 1703 AGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1762
DB 26984 CCGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27043

QY 1763 TGATGTGGGCGGAAACCGGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1822
DB 27044 TGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27103
QY 1823 AGCGGCGGCTGTTCATCTTCAATATGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1882
DB 27104 AGCGGCGGCTGTTCATCTTCAATATGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 27163
QY 1883 AGCGGAGTGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1942
DB 27164 AGCGGCGGCTGTTCATCTTCAATATGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 27223
QY 1943 TGTTCGCTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2002
DB 27224 TGTTCGCTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27283
QY 2003 TCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2062
DB 27284 GCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27343
QY 2063 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2116
DB 27344 TGTTCGAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27403
QY 2117 TGTTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2176
DB 27404 CCGTCTAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27463
QY 2177 GCGCGGAGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2236
DB 27464 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27523
QY 2237 TGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2296
DB 27524 TGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27583
QY 2297 TGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2353
DB 27584 TGTTCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27643
QY 2354 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2413
DB 27644 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27703
QY 2414 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2473
DB 27704 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27763
QY 2474 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2527
DB 27764 GGTGCTGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27823
QY 2528 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2587
DB 27824 AAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27883
QY 2588 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2647
DB 27884 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 27943
QY 2648 AGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2707
DB 27944 AGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 28003
QY 2708 CTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2767
DB 28004 TGAAGGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 28063

RESULT 13
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991

GENERAL INFORMATION:
APPLICANT: DeHoff, Bradley S.
APPLICANT: Kuhse, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sutton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match 17.7% Score 755.2; DB 2; Length 43280;
Best Local Similarity 56.0%; Pred. No. 7.8e-124;
Matches 1603; Conservative 0; Mismatches 1208; Indels 53; Gaps 7;

QY 24 CGCAGCCCAAGATCCGATTGCGATGCTGCGAGCGAGTTGCGCTGCGCCCGTGGGCTGAT 83
DB 36350 CGCGCGCGGAGCGGATGCGCTGCTGCGAGATGCGCTGCGCTAACCAGGCGGCTGTAC 36409
QY 84 CGATCTGAGCGGGTCTGAGCGCTCTCGAGGCTCGCGGCAACCGTGGGCGAGTCCC 143
DB 36410 CGGCGCCGAGAGAGCTGTGGAGCTGTGGCCGAGAGCGGAGCGGATCGGCGCTTCCC 36469
QY 144 CGCCGAGAC--GCTGGATGAGCAGCGTGGTTGATCCGACCCCGATGCCCGGAGAA 200
DB 36470 CGTGGACCGGGGCTGGAGCTGGCGTGGTGAACCGGATCCCGAGTGGAGAGGAC 36529
QY 201 GACCGCCCTTACGCCCGCATCTTCTGAGCAGCGTACCTGCTTTCGACGCTCTTCTT 260
DB 36530 CACGTAAGCCGCGGAGGCGGGTCTTGGAAAGGCGCGGTGATTCGACGCGCGCTTCTT 36589

QY 261 CGGCATCTCGCTTCGCGAAGCGCTGCGAGTGAACCTTGACATGACTCTTGTGAGAGT 320
DB 36590 CGGCATCTCGCTTCGCGAAGCGCTGCGAGTGAACCTTGACATGACTCTTGTGAGAGT 36649
QY 321 GTGCTGGAGAGCGGTGGAGAACCGCGCGATCGCTCAATCGCGCTGCTGTGGTAAAGAAC 380
DB 36650 GTCTGGAGAGCGGTGGAGAACCGCGCGATCGCTCAATCGCGCTGCTGTGGTAAAGAAC 36709
QY 381 GGAAGTTCATCGGAGATGCGGCCCGTCCGAATATGAGGCGCGCTGCGCAAGCGAGCGC 440
DB 36710 TGGTGTCTACGTGGGCGCGCGAGCGCTGTAGCGCTCGATCCCGCTGTGTGCCGA 36769
QY 441 GTCCGAGAGATGACCGCTCAATGCGGCGCTGGAGAGCATGCCAGCTGGAGCGGCGG 500
DB 36770 GGGCTCGGAGGAGCTATGCTGAGACCGGAGCGCGAGCGG-----GTAGTGTCCGCGG 36823
QY 501 AATCTCGTATGCTTGGGAGCTGCGAGGCGCTGTGTGCGGTGATACGGCTATTGCTC 560
DB 36824 CATCTCTACCGCTGCGCTGCTGCAAGAGACCGTCAATGACGGTGGAGACGGCTCTCTC 36883
QY 561 CTGCTGATGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 620
DB 36884 CTGCTGATGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36943
QY 621 CTGCTGATGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680
DB 36944 GCTGCGGAGCGGAGTGGCGGATGCGGATGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 37003
QY 681 CGGCGCTGCGCGAGAGAGCGTGTGCTGCAAGGCAATTTTCGCGAGGCGGAGCGGAGG 740
DB 37004 GAAAGGAGCTGCGCGCGAGAGCGGCTGCAAGGCGGAGCGGAGCGGAGCGGAGCGG 37063
QY 741 ACGAGCGAAGAGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 800
DB 37064 CTGCGCGAGAGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 37123
QY 801 CGATCGATATGCGCGGTGATTCGAGATTCGCGATCAATCAACGCGTGGAGCAGCGG 860
DB 37124 GCACACGATCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 37183
QY 861 TCTGACCGTGTGCGAGAGAGCTCCCAAGAAATGTGCTGAAAGGCGCTGCGGAGCGC 920
DB 37184 GCTACCGCGCGCGAGAGAGCTCCCAAGAAATGTGCTGAAAGGCGCTGCGGAGCGC 37243
QY 921 AGGCTGCGCGCGCTTCTGAGGTTATGCTGAGGAGCAGCGGAGCGGAGCGGAGCGG 980
DB 37244 CGGCTGTCTCCGAGAGAGCTGAGCGGCTGAGGAGCGGAGCGGAGCGGAGCGGAGCG 37303
QY 981 TGACCCCATGAAATTCAGAGCTTGAATGCGGTATACGCGCTCGGCGAGATGTGCGCAC 1040
DB 37304 CGACCCCATGAGAGCGGAGGCGCTGCTGCGGCTCGGAGAGGAACTTCCGCGAGCA 37363
QY 1041 GCGCTGCTGATCGGCTGCTGAGAGCAACCTTGGCATCTGATGATGCTGCGGAGT 1100
DB 37364 CCGCTGTGCTGCTGCTGCTGAGAGTCAATGCGGAGTCCGAGGCGCGCGCGGCTGT 37423
QY 1101 CACTGGGCTGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1160
DB 37424 CGGCGGCTGATCAAGATGCTTCAAGCGGCTGCGGAGCGGAGTCTGCTGCTGCTGCT 37483
QY 1161 CGCGCAGCGCTGAGACCCCGGATCTCATGAGGAGTATCTTGTGCTGACCGTCAAGCGG 1220
DB 37484 CGCGCAGCGCTGAGACCCCGGATCTCATGAGGAGTATCTTGTGCTGACCGTCAAGCGG 37543
QY 1221 CGGAGACCGTGGCGGAGCTGAGAAATAGCGCGCAGCGGCGGAGGAGTGTGCTGCTG 1280
DB 37544 CGAGTGTGCTGAGAGGAGCGGAGCGGCGCGCGCGGAGCGGAGGAGTGTGCTGCTG 37603
QY 1281 GAGGCGGAGCAAGCGGAGCGGAGTGTGAGAGAGCGGCGGAGCGGAGTGTGAGAGCGG 1340
DB 37604 CGGCGGAGCAAGCGGAGCGGAGTGTGAGAGAGCGGCGGAGCGGAGTGTGAGAGCGG 37663
QY 1341 GCGCGCGAGAGCGAC-----GGCAGAGCTGCT 1367

```
Db 37664 GCGCGGAGAGGCGCCCGGCGCTCCCGCCGCGCAGAGGGGCGGAAGGCGCCCTGCGCTG 37723
||| ||| |||
Qy 1368 GGTGCTGTGCGAGAGACCGCGCTCAGCCCTGATGACAGAGCGCGCGCTGCGGACCA 1427
||| ||| |||
Db 37724 GGTGCTGTGCGAGAGACCGCGCTCAGCCCTGATGACAGAGCGCGCGCTGCGGACCA 37783
||| ||| |||
Qy 1428 TTTGAGAGACTACCTTTCGCAATGTCTGAGCGATGTGGGCTTCACTGTGCGACGACGG 1487
||| ||| |||
Db 37784 CTTCTCCGCAACCCCGGGGCGCGCGCGCGGACATGCGCTTCTCCCTGCGCGCACCGG 37843
||| ||| |||
Qy 1488 CAGCGCGATGAGACACCGCGCTGCGGCGCGCGAGCTGAGAGGAGGCGCTGCGGACAC 1547
||| ||| |||
Db 37844 CCGACGCTTTCACACCGCGCGCTGCTGATGCGGCTGAGAGGAGCGGACCTGCGCGCGC 37903
||| ||| |||
Qy 1548 CCGTGAAGCTGCGCGCGAGAGACAGAGCTGCGCGCGCTGCGGCGGATGCGCGGATTC 1607
||| ||| |||
Db 37904 CTTGAGCGGCTGTGGCGGAAGGACGCGAGCTGCGCGGCTGTGTGCGCGGAGTCCGCGA 37960
||| ||| |||
Qy 1608 CTCAGCGCGGCAAGCTGCGCTTCTCTTTCACGGAACAGGGGGCGGACAGCGCTGCGAC 1667
||| ||| |||
Db 37961 CCGGAGCGGAGAGATGCGCTTCTCTTTCACGCGGAGGAGCGAGCGCGCGGAGTGGC 38020
||| ||| |||
Qy 1668 CCGTGGGCTGTGAGATGATGATGCGCGCTGCGGAGGCTTCACTGTGCGTGAAGCT 1727
||| ||| |||
Db 38021 CCAAGACCTGCAATGCGCGCCCAATACCTTCTTCCGCTGCGCGCTGAGAGAGTACGGA 38080
||| ||| |||
Qy 1728 GTTCAACAGAGAGCTGAGACCGCGCGCTGCGGAGGATGTGGCGGACCGGACGCT 1787
||| ||| |||
Db 38081 TCTGACCGGCTGTGCGCGCGCGCGCTGCGGCGCTGAGAGCGCGGCTGCGC 38140
||| ||| |||
Qy 1788 CAGACCGCGGCTGTGAGACCAACAGACCTTTCACCGAGCGGCGGCTTCACTTGAATA 1847
||| ||| |||
Db 38141 CCAAGCGGAGATCTGAGACCGGACGAGTACACCGAGCGGCGCTTTCGCGTCAAGT 38200
||| ||| |||
Qy 1848 TCGCGCTGCGCGCGCTGTGCGCGCTGCGGAGTGAAGCGGAGTGTGCGCGGACATAG 1907
||| ||| |||
Db 38201 GCGCGCTCCACCGCGCTGTGAGACATGCGGAGATGCGCGCGCGACCTGCTGCGGACATC 38260
||| ||| |||
Qy 1908 CATCGGAGCTGTGTGCTGCTGCTGCGCGGCGGCTTCTGCTTGAAGACCGCGGT 1967
||| ||| |||
Db 38261 GGTGCGGAGATGCGCGCGCGCGCGCGCTGCGGCTGCTGATCTGAGACGCGCTGCGC 38320
||| ||| |||
Qy 1968 CCGTGGGCTGTGCGCGCGCGCGCGCTGATGAGCGCTGCGCGCGCGCGCGGATGTGTC 2027
||| ||| |||
Db 38321 GCTGGGCGCGCGCGCGCGCGCGCTGATGAGCGCTGCGCGCGCGCGCGGATGTGTC 38380
||| ||| |||
Qy 2028 GATCGAGCGCGCGCGCGCGATGTGCTGCTGCGGCGCGCGACGACGCGTCTGCTGTC 2087
||| ||| |||
Db 38381 CCGTGGGCGCGCGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCGCGCGGAGCGCGTCTG 38440
||| ||| |||
Qy 2088 GATCGCGCGGCTGACGCTTCCGAGCCAGGTGTGCTCATCGCGCGCGCGCGGACCCGTGCA 2147
||| ||| |||
Db 38441 CCGTGGGCGCGCGCGCGCGCGCGCGCTGCGGATCTTCCGCGCGCGGAGCGGATGTGTC 38500
||| ||| |||
Qy 2148 TCGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGGACCAAGCGCGCTCAAGTCTGCGA 2207
||| ||| |||
Db 38501 CAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGGATGTGTC 38560
||| ||| |||
Qy 2208 TCGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2267
||| ||| |||
Db 38561 CCGCTTTCACCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38620
||| ||| |||
Qy 2268 GGTGAGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2324
||| ||| |||
Db 38621 CCGTGGCGCTACCGGAGAACCGGAGCTGAGCGCTGCTCAAGGTCACGCGCGCGCGCGCGCG 38680
||| ||| |||
Qy 2325 CACAGACGAGGTGAGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2384
||| ||| |||
Db 38681 CCGCGGAGATCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38740
||| ||| |||
Qy 2385 CCGGAGTGAAGTGAAGCGCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2444
||| ||| |||
```

```
Db 38741 CCGGAGCGCGGTCCGACAGCGGACACCGCGCTGAGAGCGCGGACCTTCTGAGAGACCGGCC 38800
||| ||| |||
Qy 2445 GAATCGACGCTGCTGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2504
||| ||| |||
Db 38801 GGAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 38860
||| ||| |||
Qy 2505 CGCATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2564
||| ||| |||
Db 38861 ---GCCGCGATTCACAAAGCCCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 38917
||| ||| |||
Qy 2565 CTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2624
||| ||| |||
Db 38918 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38975
||| ||| |||
Qy 2625 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2684
||| ||| |||
Db 38976 GCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39035
||| ||| |||
Qy 2685 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2744
||| ||| |||
Db 39036 CCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 39095
||| ||| |||
Qy 2745 CCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2804
||| ||| |||
Db 39096 GCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39155
||| ||| |||
Qy 2805 ACGCGGAGAGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2848
||| ||| |||
Db 39156 ATCCACCGCGGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39199
||| ||| |||

RESULT 14
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, P.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/105.537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match 17.4%; Score 742; DB 3; Length 15872;
Best Local Similarity 55.7%; Pred. No. 1,5e-121;
Matches 1594; Conservative 0; Mismatches 1215; Indels 53; Gaps 7;

Qy 26 CAGCGGAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 85
||| ||| |||
Db 8490 CCGAGGAGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 8549
||| ||| |||
Qy 86 ATCTAGCGGAGTTCGAGCGCTCTCTGAGGCGCTGCGGAGACCGTCTGCGGAGTCCCGG 145
||| ||| |||
Db 8550 CCGCGGAGAGACTGTGCGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 8609
||| ||| |||
Qy 146 CCGAAC---GCTGGAGTACAGCGGCTGTTGATCCGACCCCGATCCCGGAGGAAGA 202
||| ||| |||
Db 8610 TCAACCGCGGCTGGGACCTGAGAGACCTTACGACCCCGATCCGAGTCCGAGGACCA 8669
||| ||| |||
Qy 203 CCGCGGATACCGCGGATCTTCTCTGAGCGAGTGTGCTGCTGAGCGCTCTCTCTCTCT 262
||| ||| |||
Db 8670 CCGTCTGCGCGGAGGCGGAGTCTTCTGAGAGCGCGCGGATCTGAGCGCGCGCTCTCTCT 8729
||| ||| |||
Qy 263 GCATCTGCGCTGCGGAGCGGCTGCGGATGAGACCTGACATGACTCTTGTGAGAGTGT 322
||| ||| |||
```


Db	8730	GCATTCGCGCGCGAGGCGCTTGTAATGACCCGACAGCGGCTGCTGTAAGGTG	8789
Oy	323	GCTGGAGAGCCCTGGAGAACGCCGCGATCGGCTTCATTCGCGCCTCGTCGTAACGGAACG	382
Db	8790	CCTGGAGAGCCCTGGAACGCCCGCGCATTCACCCGTCCTCGTCGCGCGACGCCGCGGT	8849
Oy	383	GAGTGTTCATCGGAGTCGCGCCGTCCTCGAATATGAGCCGCGCTGCGACGACGCGGT	442
Db	8850	GTCATACGTGGCGCCCGACCGACTGTAAGCTCCGATCCCGGCTGGTGCCGAG	8909
Oy	443	CCGAGAAATGACGCTATAGCGCGGCTGGGGAAGATCCAGCGTCGAGCGGCGGAA	502
Db	8910	GCTCGAA-----GCTATCTGTGACCGGACGCGACGCGGTGATGTCGCGCGCA	8963
Oy	503	TCTGCTATGCCCTCGGCGTGGAGGCGCGGTGTGCGCGTGAATACGACTATTGTCCT	562
Db	8964	TCTCTAAGCGCTCGGTCTCGAAGGACCGTCATATACGTATGAGACGCGCTGCTCTCT	9023
Oy	563	CGTGTGCGCGTTCATTCGCGCTGTCAAGCTTCGTCGCGGAAATGCTTCACGCGCC	622
Db	9024	CGTGTGTGCGCTGATCTGGCGGTACGCGCGCTTCGCGACGCGAGTGGCGGCTCGCG	9083
Oy	623	TGGCTGTGGGGTATTCGCTGATTTGTGTGCGCGAGACCCCTGTGTGCTTCGAAGACC	682
Db	9084	TGGCGGGGGGGGTGGCGGTGATGCGCATTCGCGCGCGCTTGTGTGAATTCCTCCGCGAGA	9143
Oy	683	GCGCGCTGGCGAGGAGCGGTGCTTCAGAGCATTTTCGCGGAGGCGCATGAGGTTCGAGC	742
Db	9144	AGGGGCTGCGCGCGACGCGCGGCTTCAGAGCGTTTCGCGCGCGCGCACGCGACCGCT	9203
Oy	743	GAGCGGAAGGTGCGCGCTGTGTCTTCAGACGCGCTCATGTGAAGCCCGCGCGAGCGCG	802
Db	9204	GGGCGGAAGGCGTGGCGTGTCTGTCTGTGAAGCGGCTGTGCGACGCGCGCGCGGCGGCG	9263
Oy	803	ATCGGATATTGGCGGTGATTGAGGATTCGCGATTCATACAGACGCTGTGAGCAGCGCTC	862
Db	9264	ACACGCTCTCGGCGTGTCAACCGGACCGCGGTCAACAGAGCGTGTCTTCACACGCGC	9323
Oy	863	TGACCGTCCGAAACGAGAGCTCCCAAGAAATGATGTGTAAGACGCGCGCTGGCGAGCGAG	922
Db	9324	TGACCGCGCCCAACGCGCCAGCCACGACATCGGTATGCGCGAGCGCTTCGCGACGCG	9383
Oy	923	GCTGCGCGGCTTTCGGTGAGTTATGTGAAGGACACGCGACGCGGACGACGCTTGTG	982
Db	9384	GGCTGTCCCCGGAAGACGTGAGCGCGGTGAGGCGCACGCGACCGGACCCGCGCTCGCG	9443
Oy	983	ACCCATGAAATCCAACTCTGAATGCGGTATACGCGCTCGGCGGAGATGTCCGACGC	1042
Db	9444	ACCCATGAGAGCCGCGCGCGCTGTGCGCGCTCCGACGGAACGTTTCGCGCGACAC	9503
Oy	1043	CGCTGCTATGGGTCGCGTGAAGACCAACTTGGGCATCTGAGATAGCTCGGGATCA	1102
Db	9504	CGCTGTGCTGCGCTCGCTGAAGTCCAAATCGGGACATGCCCCAGCGCGCGCGGTGTG	9563
Oy	1103	CTGGGCTCTGAAGGTGCTTGTTCCTTCAGACGAGGACAGATTCTTCGCACTTCAGC	1162
Db	9564	GCGGCTCATCAAGATGCTCCAGGCGCTCGCGACGCGTGTGCTCCGCGCACCTTCACG	9622
Oy	1163	CGCAGCGCTGAACCCCGGATCTCATGGGTGATCTTCGCTGACGTCACGCGGCC	1222
Db	9624	CCGACGACCGACCCCGCATGCGACTGAGGCTCGGCGCGGTACGCGTCACTTCG	9683
Oy	1223	GGAACCGGTGCGGACCTGGAATACGCCCGACGCGCGGGGTGAGCTGTTCCGCAATGA	1283
Db	9684	AGTGTCCGTGGAGGGAACCGCGCGGCGCGCGGATGTGCGGCTTCGCGCTGCG	9743
Oy	1283	GCGGGAACCAACGCGACGTTGTGTGGAAGAGCGCGCGCGCGAGTGCACACGCGCG	1342
Db	9744	GCGGGAACCAATGCGCATGTGCTTCGAAGAGGACCGCGCGCGCGCGCGGAACCG	9803
Oy	1343	GCGCGAGCGAAC-----GCGAAGCTGCTGG	1365

Db	9804	CCGGGAGAGGGCCCCCGGCGCTTC	CCCGCGCCGAGAAAGGGCCGAAAGGGCCCTCGGCTGGG	9863	
OY	1370	TGCTGTGGCAAGAACCCGCTCA	CGCCCTTGATGCAACAGCGCGCGCTGGCGACATC	1429	
Db	9864	TGTCCTCGAGCGGAA	CGAGCCGGCCCTGGGATCCAGAGCCCGCGCTCCCGACAC	9923.	
OY	1430	TGAGACCTTACCTTCGACATGT	CTGGGCGATGTGGCGTTCAATCTGGCGACAGCCGA	1489	
Db	9924	TTCTCCGACACCCCGGGGCCCG	CGCGCTGACATCGCCCTTCCCTCGCGCGCACGCGG	9983	
OY	1490	GCGCGATGAGACACCGGCTCG	CGGTTGGCGGACGTCGAGGAGGGAGCTGGCGGACCC	1549	
Db	9984	CAGCTTTTGAACACCGGCGCC	GTCGTATCGGCTCGAAGGGGCCGAATCGCCCGCC	10043	
OY	1550	TGAGACGTGCGGCGGAGGAC	AGACGTGCGCCGATGCGGTGGACGATTCGCAATTC	1609	
Db	10044	TGAGACGCTTGAGCGGAGAG	ACCGACGCGATCGGTGGCGGAGTCGCGGATCCGCA---CC	10100	
OY	1610	CACGCGGCAACTGGCTTTTCT	TTTACCCGACAGGGGGCGGACGCTGGGCAATGGGC	1669	
Db	10101	GAGACGGGAGATGGCTTTC	CTTTCACCCGAGCAGGAGACGCGCCCGGATGGCC	10160	
OY	1670	GTTGGCTGTACGATATGAT	GTCCGCGTCCGCGAGGCGCTTCGACCTGTGCTGAGGCTGT	1729	
Db	10161	ACGACTTCATATGCCGCCAT	ATACCTTTTCCGCTCCGCCCTCGACAGATGACGACCTTC	10220	
OY	1730	TCAAACGAGACTCGACCGG	CCGCTCCGCAAGTATGTGGCCGAACCGGCAAGGTCG	1789	
Db	10221	TGACACCGCTCTGGCGCGG	CGCTCGGCGGCTGTGGAGCGCCGACCCGAGCTCGCCG	10280	
OY	1790	ACGCCGCGCTGCTGACCA	AGACGCTTCAACCAGCCGCGCTTTACCTTCGAATAG	1849	
Db	10281	AAGCGGCACTCTTGACCG	GACCGGATCACCCAGCCGCGCTCTTCCGCTCGTAGAGTGG	10340	
OY	1850	GCGTCGCGGCGCTGTGGG	GCTGTGGGATGAGACCCGAGTGTGGCCCGGCGCAATAGA	1909	
Db	10341	GCGTCACCGGCTGTGGAG	ACATGGGGGATGCGCCCGGACCTGCTCTGGGCACTCGG	10400	
OY	1910	TGCGTGAAGCTGTGTGCT	CTTCCGTGTGGCGGCGTTCCTCGCTTAGAGA	CGCGATGTTCC	1969
Db	10401	TGGGGGAACTGGCGCGG	CGGCCACGTCGGGGGTGTGATTCGATTCGACGACGCTCGCGG	10460	
OY	1970	TGTGTGCTGCGCGCGGCG	CGCTGATGACGCGCTCCCGCGCGCGGCGGAGATGTGTGA	2029	
Db	10461	TGTGTGCGCGCGCGGCA	GGCTGATGACGCGCTCCCGCGCGCGGCGGAGTGTCTCG	10520	
OY	2030	TGAGAGCGCCCGAGCGG	CGATGTGGCTGTGGGTGGCGCCGCAACGACGTGTGGTGA	2089	
Db	10521	TGCGGGCGGCGAGAGAC	AGAGTCCGCGCACTGTCTGCGCGCGCGGCGGACGCTGTGCG	10580	
OY	2090	TGCGCGCGATCAACGCT	CCGGAACAGTGTGATGCGCGGCGCCCGGCAACCCGTGTATG	2149	
Db	10581	TGCGCGCGGTAAAGGCCC	CGCGGTGGTGTATTTCCGCGCGGAGAGAACGCTGTGGCG	10640	
OY	2150	CGATGCGCGGCGGATGG	CCGCGCGCGGCGCGGAGACCAAGGCGCTTCAGCTTCGCAATG	2209	
Db	10641	AGGCGGCGGCGGAGCT	CGCCGGAAGGACCGGCGCGGCAACGAGCGGCTCCGCTCGCGACG	10700	
OY	2210	CGTTTCACTACCGCTCA	TGGCCCCGATGTGTGAAGCGTTGGGCGGTGTGGCGATCGG	2269	
Db	10701	CCTTCCACTCAACCTTGA	TGAGCGCAATGTCCGCGAATTCGCGGAGGTGCGCCCGCGGCG	10760	
OY	2270	TGAGCTACCGGCGGCGGT	CGATTCGCTGTGTCAGCAATTCGAGCGGAGG---GCTTGA	2326	
Db	10761	TGCGTTACCGGAGACCG	AGCTGACGTCGTTCACAGGTACGAGGCGGCCCCCGCCG	10820	
OY	2337	CAGACGAGGTAGCTGCG	CGGCGCTATTGGGTGCGCACGCGCGGAGAGGTGTGCGCTTG	2386	
Db	10821	CGGATGAATCAACCGG	CCCCCGACTATCTGGGTGGCCAGATCCGTGAGCCCGTTCG	10880	
OY	2387	CGGATGAGTGAAGCGCT	GCAACGCGGCGGCTGTGGGCACTTCTGTCAAGGTTCGATCGA	2446	
Db	10881	CGGACGCGGTTCGCA	CGGCAACCGCTCTGAGGCCCGACCTTCTGTGAGACCGGCGCG	10940	

Qy	2447	AATGACGCGTCTCGGCTGTGGTGTGCTCGTGAATGCGGACCGCGCGGCGGTGTGCG	2506
Db	10941	ACGGCGTGTGTGTGGGATGATGACAGAGAGATGCTTGAGAGACACACCTGTGGCTTGTCC	11000
Qy	2507	CATGTCGCGCGCTGGGCGTGAAGAGCGGACGCTGTCAAGGCGCTCGGCGGCTCT	2566
Db	11001	---CGGGATTCACAAACCCGGGACCGCGCGGACGATCCGGGGCTCCGGGGCGTGC	11057
Qy	2567	GGGCGGTGTGGTGTGTGTCTCTGTGGCGGGCTTTCCCTCAGGGGGCGCGGATGC	2626
Db	11058	GGGGGGCGCGCGCGCGCTTACGAGCGGGGGCGCCCGGGGTGACT--GGGCGGGATGACCGC	11115
Qy	2627	CGCTGGCCAGGTACCTTTGGACGGGAGGCGCTACTGAATGCACAGAAAGCGACGAG	2686
Db	11116	CGACGGCCCGGAGGGGCGGCGCGCGCGCTGTGAATCTGCCCCGTCAAGCCTTCGGACCG	11175
Qy	2687	CGGCGCTGTGGACCGCGCGTCTCCGGAGCGGGTCAAGACGAGTTCGAGAGGGGGGCG	2746
Db	11176	CCGCTACTGGCTGGCCCGGGCGCGCGCGGACACAGACATGTGATGTACCGGATCGG	11235
Qy	2747	CGGTGCCCGGCGGACCGCGCGCTGTGGCTGCACATCCGCGCGCCGAGAGCGGAC	2806
Db	11236	CTGGGACCGGCTGTGGCTGTGACCGGCGGGGCGCGGACCGCGCGCGCTGTGTGAT	11295
Qy	2807	GCCGGAGAAAGTGAAGCGCGCGGCGGACCGTCTTCCGCGC	2848
Db	11296	CCACCCGACAGCCCGCGCGTGTCCGGGAGCTGTCCGGCGACGC	11337

RESULT 15
US-09-103

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A

```

? GENERAL INFORMATION:
? APPLICANT: FLEISCHMAN, Robert D.
? APPLICANT: WHITE, Owen R.
? APPLICANT: PRASER, Claire M.
? APPLICANT: VENTER, John C.
? TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
? FILE REFERENCE: 24366-20007.00
? CURRENT APPLICATION NUMBER: US/09/103.840A
? CURRENT FILING DATE: 1998-06-24
? NUMBER OF SEQ ID NOS: 2
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 1
? LENGTH: 4411529
? TYPE: DNA
? ORGANISM: Mycobacterium tuberculosis
? OTHER INFORMATION: H37Rv
? JS-09-103-840A-1

```

Query Match	17.4%;	Score	740.8;	DB	3;	length	4411529;	
Best Local Similarity	56.7%;	Pred.	No. 3.3e-121;					
Matches 1571;	Conservative	0;	Mismatches	1127;	Indels	74;	Gaps	8

QY	19	AAGGCGCAGCCGAAGATCCGATTGCCATCGTCGGAGCGAATTGGCCGTCGCCGGTGGC	78
Db	3297714	GAGCAGACCGCGCAGACGAAACCGGTGGCGGTGTGGGAAATCGCGCTGCCATTTTCCGGGCGGC	32976555
QY	79	GTGATCGATCTGAGCGGGTCTTGAGACGCTCCTCGAGGGCTCGCGCGACACCGTCGGGCGA	138
Db	3297654	GTGATGAGCCCGGACGGAAGACTCTGGGAGATGTCTTTCCGGCGGGGCGCGAAGTGTGTGCGGAG	32975959
QY	139	GTCCCCCGCCGAAC--GCTGGGATGCGACGACGTCGTGTTGATCCCGAACCCCGATGCCCGC	195
Db	3297594	TTTCCGAGCCGATCGGGGCTGGGAGTGTGAGGGGCTAATCGATCCCGAATCCCGAAGCTGAG	32975359
QY	196	GGGAAGAGCGCCGTTAGCGCGCGCATCTTCTCGAGCGCAGCAGTAGCCTGTTGACGCGCTCC	255
Db	3297334	GGCAAAACTTACAGCGCATGGGGGGCATTTTCTAGACGCGCGACCGGCTTTGAGCGCCGGG	32974757

OY	256	TTCTTGGCAATCTCGCTCCCGAAGCCGCTCGATGAGACCTGCACATTCACACTTCCTG	315
Db	3297474	TTCTTCGGCAATCGCTCTTAATGAAGTGTTCGACATGGAATCTCTCAGCAGCCGGTGAATCTG	3297415
OY	316	GAGGTGTGCTGGAGGCGCTGAGAAAGCCGCGAATGCTCCATGCGCGCTCTGCGTAGC	375
Db	3297414	GAGGTTCGTGTGGAGGCGTTGGAGCAGCCGGGGAATTGAACCCGTGTGTGTCGCGTGTTCG	3297355
OY	376	GAAACGGGAATGTTTCATTCGGGAATGCGCCGCTCCGAATATATAGGCCGGGCTGCGCAAGC	435
Db	3297354	GCGACCGGGGTGTATACACCGGCATTTTTCGCGCGAGTTACG-----GCAAT	3297310
OY	436	ACGGCGTCCCGAGAGATCGACGCTCATGCGCGGCTGGGGACGAATGCCAGCGTCCAGCG	495
Db	3297309	AGGGATATCCGGGGGGGCTGCAGAGGATACGGGTTGACCGGACGTCATATCAGCTGACCTCG	3297250
OY	496	GCGCGAATCTCTGATATGCGCTTCGCGGCTGCGAGGGCCGTGTGTGCGCGTGAATACGGCTAT	555
Db	3297249	GGAAGTGTGTGTATCGATATTTGGGGGCTCAGAGGCCCGCGGGTGTACATGAGATACCGTGTG	3297190
OY	556	TGCTCTTCGCTGGTGGCCGTTTCATCTGGCCTGTACAGCTTTGGGCTTCGGGGAAATGCTTC	615
Db	3297189	TGCTGCTGTGTGGGTGGGATCAATGGCGCATGTGTCATATGGCGCTTCGGGGAGTGTGAT	3297130
OY	616	ACGGCCCTGCGTGGGGGTATCGCGTATGTGTGCGCGAGCAACCCCTGATGGCTCTCG	675
Db	3297129	TTGGCGTTTGGCGGGTGGGGGTACGATATAGGCGTGGCCGTGCATTTTGTGTGGGTTTTTC	3297070
OY	676	AAGACCCGGGCGCTGCGCAGAGGACGATCGCTGACAGGCAATTTTCGCGAGGCGCGATAGG	735
Db	3297069	CGGACGGGGGGTTGGCTGCGCATAGGCGCATATGAAGCGTTTGCGCGCGGCGGACAGCG	3297010
OY	736	TTTCGACGAGGCGAAGGGTTCGCGCTGTGTCTCTCAAGCGGCTCAATGAGACCCCGCGC	795
Db	3297009	ACCGGATGGGCGAGGCGCGCGGGTGTGTGTCTGAGAGCGGCTATCGACGCGCCGCGG	3296950
OY	796	GACGCGCATATGGAATTTGGGGGTGTATTCAGAGATCCGCGATCATACAGACGGTGGAGC	855
Db	3296949	TTGGGAGCATTCGAGTGTGGGCTGGGTGTGTCGGGGAGCGCGGTCAACAGAGATGGCTCTC	3296890
OY	856	AGCGGTCTGACCCGTCGCGAAGCGGAGCTCCCAAGAAATCTGTCTGAACAGGCGCTTGGC	915
Db	3296889	AATGGGCTGACCGGCGCCCATATGGGCTGCGCCACGAGGAAATTCAGGTGTGGTTGGC	3296830
OY	916	GACGCAAGGCTGCGCGCGCTTTCGATGGGTTTATGTGAGGACACAGGCAAGGCAAGC	975
Db	3296829	AACGCGGGGTGTGTCCGCGCAGACGTAGAGTGGTGGAGGCAACAGGACCGGCAACAGC	3296770
OY	976	CTTGTGTATCCCATTCGAAATTCGAAAGCTGTGAATGCGGTATATACGGCCTCGGGCGAAGATC	1035
Db	3296769	TTGGGTATATCCCATCGAAGCGCAGGCAATTCGTCTCATCTTACGGGCAAGGAGTCCAGCG	3296710
OY	1036	GCCACGCGCGCTGCTGATCGGGTGTGTATAMACCAACCTTGGGCATCTCTGATATGCTCG	1095
Db	3296709	GAGACGCGCTGTGTGGGTGGGCTCAATCAATCCAACTGGGCGACACCCAGCGCGCGCGC	3296650
OY	1096	GGAATCACTGGGCTGTGAAGTGTCTTGTCTCTTCAGACAGGCGAGATTCCTGCGCAC	1155
Db	3296649	GGGGTCCCGGGGTGATCAAGATGTGTACAGGCGATGCTCATGGGAGTATGCGGCGAGC	3296590
OY	1156	CTCCACGCGCAGGCGCTGMAACCCCGGATCTCATTGGGTATCTTTCGGCTGACCTGTACG	1215
Db	3296589	TTGATATGTGATGAGCTTTCCTCCCGGGGTGATTTGACACAGCGCGCGGTGTGTGTTTG	3296530
OY	1216	CGGCGCCGGACACCGTGTGGCGGACTGMAATATACGCGCGAGCGGGCGGGGTGAGCTGCTTC	1275
Db	3296529	ACCGAGGCCAGGAGTGTGTGATTGACGGCGCTTCGCGCGGGCGCGGTGTCTCTCTTT	3296470
OY	1276	GGCATATGAGCGGACCAACGCGCACGTGTGTCTGTGAAGAGCGCGCGCGAGC-----	1328
Db	3296469	GGGATTCAGCGGCAACACGCGCATCTGTATCTTGAATAGAGGCCCTCCGTATACAGGCCCGCA	3296410
OY	1329	-----GTGCACACCGCGGCGCGCGCGAGCGACCGGCAAGACTGCTGG	1369

Dh 3296409 GAGCGCGCGTTGAGCGAGGAAAGCACCGGGGGCCGCGCCCTGATGTGCGGTGG 3296350
Qy 1370 TGCTGTGGCAAGAGACCGCGCTCAGCCCTGATGACAGAGCGCGCGCTGCGCAATC 1429
Dh 3296439 TGATTTGGCGAGATCGCTGAGGCGTTGACGGCGAGCGCGGCTGACTTATGCGCA 3296290
Qy 1430 TGAAGACCTACCTTTCGAGTGTCTGAGCGATGTGGCGTTGACTGTGCGACGCGCA 1489
Dh 3296289 TGCAGCGCAACCCAGGGCTGGATCCGATCGATGTGGGTGTCTGTTGGCCAGT---CGCT 3296233
Qy 1490 GCGGATGAGACACCGGCTGCGGTTGGCGGCGACGTGAGGAGGAGGTTGCGGACGCC 1549
Dh 3296232 CGGTGTTGAGCACCGAGCGGTGTGTGCGCGCAAGCCGTGACCACTGATTCGCGGC 3296173
Qy 1550 TGAAGCTGCGCGCCAGGAGACAGAGTGCACCGGTGCGGTGCGAGTATCGCCGATTCCT 1609
Dh 3296172 TGCTGAGGCTGCGCGCGCGCGAGCGGGTGCAGCGTGTGCGGTGATGAGCCAG--GCT 3296116
Qy 1610 CACGCGGCAAGCTTCCTTCTCTTCAACGAGAGGGGCGCAGACGCTGCGCATGAGGCC 1669
Dh 3296115 CGGTGCGCAAGACGCTGCTGTCTGTTCTTGGCGAGGGGCGCAGCGCATCGGATGAGGCC 3296056
Qy 1670 GTGGCTGTGACATGATGTGTCCCGCTTCCGAGAGCGCTTGCACCTGTGCGTGAAGCTGT 1729
Dh 3296055 GCGAGTGTATCGGCGAGTGTCCCGTGTGCGCAGGCAATGATGCGGTGCGCAGAGT 3295996
Qy 1730 TCAACGAGAGCTGACCGCGCGGCTCGGAGGATGTGGGCGGAAACCGCCAGCGCTCG 1789
Dh 3295995 TGAACCGGCACTTCGCGTTCGCGCTGCGCGAGCTTATTTGGG-----GTGCGG 3295948
Qy 1790 ACGCGCGCTGCTGACACAGACACCTTACCCAGCGCGCTGTTCACCTTCCGATATG 1849
Dh 3295947 ATGCGGATTTGCTTGACAGACCGAATTTGCTCAGCCCGCGTGTGCGCGGTGAGGTTG 3295888
Qy 1850 GCGTCGCGCGCTGTGTGCGGTGTGTGGGTGTAGACCGGAGTTGTCGCGCCCATAGCA 1909
Dh 3295887 CATGCTTCGCGGTGTTCGCGGATTTGGGGTGTGCTTCGCGACTTCATGTGCTCATCCG 3295828
Qy 1910 TCGGTGAGCTGTGCTGCTGCGGAGCGCGGTCTGCTGCTGAGAGCGCGGTTC 1969
Dh 3295827 TTGAGAGAGCTGCGCGCGCGCGACCGCGCGGTGTGTGATCTTGCGCGAGCGCGGATGC 3295768
Qy 1970 TGTGTGCTGCGCGCGCGCGCTGATGACAGCGCTGCGCGCGCGCGCGATGTGTGA 2029
Dh 3295767 TGTGTGTGCGCGCGCGCGCGGTGATGACAGCGCTGCGCGCGCGCGCGATGTGTGA 3295708
Qy 2030 TCGAGCGCGCGAGCGGATGTGCTGCTGCGGTGCGCGCGCGCGCGAGCTGCTGCA 2089
Dh 3295707 TGGCTGCGAGTGAAGACGAGGT-----GAGCGCGCTGCTGAGGTGAGGTGTGGGA 3295657
Qy 2090 TCGCGCGGCTGCAAGCTTCGCGGACAGAGGTGATTCGCGCGCGCGCGGCAACCGCTGATG 2149
Dh 3295656 TCGCTGCGATCAAGCGCGCGCAATCGGTGTGATCTCGGTGCGCGAGCGCGCGCAATG 3295597
Qy 2150 CGATGCGCGCGCGATGCGCGCGCGCGCGCGCGCAACCAAGCGCTTCACTGCTGATG 2209
Dh 3295596 CGATTGCGGATCGGTTCGCGCGCGAGGTTCGCGCGGTGCAACAGTTGCGGCTTCGATG 3295537
Qy 2210 CGTTCACTACCGCTCATGCGCCCGATGCTGAGAGCGTTGCGCGGTGTGCGCGAGTCCG 2269
Dh 3295536 CGTTTCATTGCGCGCTGATGAGCGGATGCTGAGAGGTTCGCGGTGTGCGCGCGCG 3295477
Qy 2270 TGAAGTACCGCGCGCGCTGATTCGCTGCTGACGAATCTGAGGCGGAGAGCTTGCACAG 2329
Dh 3295476 TGCAGGCAACCGAGCCCGAGCTTGTGCTGTGTGATCGAAGCTGACGAGTGTGCGCGCC 3295417
Qy 2330 ACGAGTGAAGCTTCGCGCGGCTATTTGAGTGCAGCGCGCGAGAGTGTGCTTTCGCG 2389
Dh 3295416 CTGATTTGCGGTTCGCGCGAGTACGAGTGTGAGCAAGTTGCTGCGCGCGCTTTCGCG 3295357
Qy 2390 ATGAGTGAAGGCGCTGACCGCGCGGTGCGCGCACTTGTGAGGTGCTTCGAAT 2449

Dh 3295356 ACAGTGCGGCTATTTGACAGACCTTGGGGGAGACCACTTCAATCAGAGCGCGCGGAA 3295297
Qy 2450 CGAGCTGCTCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2509
Dh 3295296 GTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3295237
Qy 2510 GCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2569
Dh 3295236 TGTGCGGCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3295177
Qy 2570 CCGTGGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2626
Dh 3295176 CCAGCGGTGTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3295117
Qy 2627 CGTGTGCGGATACCTTGTGCGAGCGCGCTTACTGATGACACGAAAGCCGACGACG 2686
Dh 3295116 AGCTGCCAGTATGCTTTCAGGAGAGCGCGCTTGTGGAGAGCGCGCGCGGATGAG 3295057
Qy 2687 CCGCGGTGCGAGCGCGCGCTGCTGCTGCGGAGCGGCTGACAGAGTGAAGAGGCGGCG 2746
Dh 3295056 CCGCGATGCGCGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3294997
Qy 2747 CCGTGGCGCGCG 2758
Dh 3294996 TCGAGCGCGCG 3294985

Search completed: October 4, 2003, 03:39:32
Job time : 275.999 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:29:26 ; Search time 1043.88 Seconds
(without alignment)
11031.749 Million cell updates/sec

Title: US-10-014-717-1_COPY_7610_11875

Perfect score: 4266

Sequence: 1 gtggcgatgcgtccatcga.....gggaatcattgcctatga 4266

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4266	100.0	68750	21	AAZ55887
2	4187.6	98.2	71989	21	AAA29349
3	817.2	19.2	67459	25	ABO77491
4	814.2	19.1	3978	21	AAZ55785
5	814.2	19.1	12349	21	AAZ55840
6	814.2	19.1	18331	21	AAZ55857
7	805	18.9	27541	22	AAD17185
8	805	18.9	125401	22	AAD17186

9	803.2	18.8	33529	23	AAZ17367
10	766.6	18.0	20394	22	AAZ24892
11	755.2	17.7	43280	18	AAZ60413
12	745.8	17.5	53789	19	AAZ71187
13	742	17.4	15872	21	AAZ87283
14	740.8	17.4	4411529	22	AAZ96882
15	735.6	17.2	15872	18	AAZ68715
16	732	17.0	4403765	22	AAZ96883
17	726.2	17.0	65140	22	AAZ17184
18	725	17.0	49377	19	AAZ05287
19	701.2	16.4	16767	22	AAZ88339
20	701.2	16.4	50000	22	AAZ88313
21	679.2	15.9	80161	20	AAZ21501
22	679.2	15.9	14775	22	AAZ68338
23	679.2	15.9	50000	22	AAZ68336
24	677.2	15.9	75235	25	AAZ55557
25	675.2	15.8	44377	18	AAZ78558
26	675.2	15.8	44377	18	AAZ80414
27	671.8	15.7	28598	17	AAZ06769
28	671.8	15.7	28958	18	AAZ89956
29	671.8	15.7	28958	21	AAZ75239
30	671	15.7	9513	22	AAZ88337
31	663.2	15.5	29879	14	AAZ06806
32	662.2	15.5	13987	18	AAZ80415
33	649.4	15.2	47981	22	AAZ30757
34	639	15.0	6459	22	AAZ68336
35	633.2	14.8	5088	22	AAZ69033
36	633.2	14.8	34071	22	AAZ69033
37	633.2	14.8	42717	22	AAZ69032
38	633	14.8	77536	21	AAZ14651
39	628.8	14.7	4403765	22	AAZ96883
40	628.8	14.7	4411529	22	AAZ96882
41	627.6	14.7	13842	21	AAZ87297
42	627.6	14.7	36778	21	AAZ87318
43	627.6	14.7	37948	21	AAZ87285
44	627.6	14.7	38506	21	AAZ15633
45	627.6	14.7	38506	21	AAZ56001

ALIGNMENTS

AAZ55887	standard; DNA; 68750 BP.
AAZ55887	10-APR-2000 (first entry)
AAZ55887	Sorangium cellulosum 68.75 kb contig.
AAZ55887	Epithelone biosynthesis; type I polyketide synthase; taxol substructure; anticancer; ds.
AAZ55887	Sorangium cellulosum.
AAZ55887	Location/Qualifiers
AAZ55887	1..1826
AAZ55887	/*tag= a
AAZ55887	/product= "Partial Orf 1 protein (AAZ55880)"
AAZ55887	/note= "No initiation codon given in the specification"
AAZ55887	complement (1900..3171)
AAZ55887	/*tag= b
AAZ55887	/product= "Orf 2 protein (AAZ55881)"
AAZ55887	3415..5556
AAZ55887	/*tag= c
AAZ55887	/product= "Orf 3 protein (AAZ55882)"
AAZ55887	complement (5612..5992)
AAZ55887	/*tag= d
AAZ55887	/product= "Orf 4 protein (AAZ55883)"
AAZ55887	6226..6675

DNA sequence of S.
Pimaricin biosynth
Tylactone synthase
Amycolatopsis medi
S. venezuelae vep
Mycobacterium tube
Streptomyces venez
Mycobacterium tube
Streptomyces nours
The soraphen biosy
S. spinosa DNA fra
S. spinosa DNA fra
DNA fragment of Sa
S. spinosa DNA fra
S. spinosa DNA fra
Saccharopolyspora
Platenolide syntha
Sorangium cellulos
DNA sequence of So
S. spinosa DNA fra
eryA region of S.
Hybrid smg/tylg O
Micromonospora meg
S. spinosa DNA fra
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Nucleotide sequenc
Mycobacterium tube
Mycobacterium tube
S. venezuelae macr
S. venezuelae pik
S. venezuelae pik
Nucleotide sequenc
Recombinant cosmid

```
FT      /tag= e
FT      /product= "Orf 5 protein (AAV58584)"
FT      /tag= f
FT      /product= "Type I polyketide synthase, EPOS A
FT      (AAV58573)"
FT      11872..116104
FT      /tag= g
FT      /product= "Non-ribosomal peptide synthetase, EPOS P
FT      (AAV58574)"
FT      16251..21749
FT      /tag= h
FT      /product= "Type I polyketide synthase, EPOS B
FT      (AAV58575)"
FT      21746..43519
FT      /tag= i
FT      /product= "Type I polyketide synthase, EPOS C
FT      (AAV58576)"
FT      43524..54920
FT      /tag= j
FT      /product= "Type I polyketide synthase, EPOS D
FT      (AAV58577)"
FT      54935..62254
FT      /tag= k
FT      /product= "Type I polyketide synthase, EPOS E
FT      (AAV58578)"
FT      62369..63628
FT      /tag= l
FT      /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT      (AAV58579)"
FT      63779..64333
FT      /tag= m
FT      /product= "Orf 6 protein (AAV58585)"
FT      complement (63853..64290)
FT      /tag= n
FT      /product= "Orf 7 protein (AAV58586)"
FT      64363..64920
FT      /tag= o
FT      /product= "Orf 8 protein (AAV58587)"
FT      complement (64287..64727)
FT      /tag= p
FT      /product= "Orf 9 protein (AAV58588)"
FT      65063..65767
FT      /tag= q
FT      /product= "Orf 10 protein"
FT      complement (65008..65874)
FT      /tag= r
FT      /product= "Orf 11 protein (AAV58590)"
FT      complement (65871..66338)
FT      /tag= s
FT      /product= "Orf 12 protein (AAV58591)"
FT      66667..67137
FT      /tag= t
FT      /product= "Orf 13 protein (AAV58592)"
FT      67334..68251
FT      /tag= u
FT      /product= "Orf 14 protein (AAV58593)"
FT      68346..68750
FT      /tag= v
FT      /product= "Partial Orf 15 protein (AAV58594)"
FT      /note= "No termination codon given in the specification"
XX      W09966028-A2.
XX      PD
XX      23-DEC-1999.
XX      PF
XX      16-JUN-1999; 99WO-EP04171.
XX      PR
XX      18-JUN-1998; 98US-0099504.
XX      24-SEP-1998; 98US-0101631.
XX      PR
XX      05-FEB-1999; 99US-0118906.
```

```
PA      (NOVS ) NOVARTIS AG.
PA      (NOVS ) NOVARTIS-ERFINDUNGEN VERN GES MBH.
XX      Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D,
XX      WPI; 2000-097741/08.
DR      P-PSDB; AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
DR      AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
DR      AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
DR      AAV58592, AAV58593, AAV58594.
PT      New isolated epoethione synthase genes, used for the recombinant
PT      production of epoethione for use in cancer therapy
XX      Claim 14; Page 87-104; 174pp; English.
XX      This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX      comprising 22 open reading frames (ORFs) and includes genes encoding
XX      proteins involved in the biosynthesis of epoethiones. Epoethiones A and
XX      B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX      starter unit; polyketides being synthesised from two-carbon building
XX      blocks, the beta-carbon of which always carries a keto group. Each round
XX      of two-carbon addition is carried out by a complex of enzymes known as
XX      the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX      EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX      the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
XX      and EPOS E (AAV58575-158578) are involved in polyketide backbone
XX      formation. EPO F (AAV58579) is an epoethione macrolactone oxidase, and
XX      the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX      involved in transport. Epoethiones mimic the biological activity of
XX      taxol, and may be substituted for taxol in cancer chemotherapeutic
XX      compositions. Epoethiones exhibit a much lower drop in potency against a
XX      multiply drug-resistant cell line compared with taxol, and are
XX      considerably less efficiently exported from such cells by the multidrug
XX      resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX      epoethiones as anticancer agents, they are problematical to produce on a
XX      large scale. Epoethiones are too complex for industrial scale chemical
XX      synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX      poor yields of epoethiones. The nucleic acids of the invention may be
XX      used for the recombinant production of epoethiones in a heterologous host
XX      that is more amenable to fermentation.
SQ      Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;
Query Match 100.0%; Score 4266; DB 21; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 GTGGCGGATCGATCCCATCGAGCGCGACCGAAGATCGATTGCGATGCGAGCGAGT 60
DB      7610 GTGGCGGATCGATCCCATCGAGCGCGACCGAAGATCGATTGCGATGCGAGCGAGT 7669
QY      61 TCCCGTCTGCCCGATCGGCGTGAATCGATGAGCGGGTTCTGAGCGCTCTCGAGGGCTCG 120
DB      7670 TCCCGTCTGCCCGATCGGCGTGAATCGATGAGCGGGTTCTGAGCGCTCTCGAGGGCTCG 7729
QY      121 CGCGACACCGTGGGGCGAGTCCCGCGAAGCGTGGATGCGAGCGCGTGGTTGATCC 180
DB      7730 CGCGACACCGTGGGGCGAGTCCCGCGAAGCGTGGATGCGAGCGCGTGGTTGATCC 7789
QY      7730 CGCGACACCGTGGGGCGAGTCCCGCGAAGCGTGGATGCGAGCGCGTGGTTGATCC 7789
DB      7730 CGCGACACCGTGGGGCGAGTCCCGCGAAGCGTGGATGCGAGCGCGTGGTTGATCC 7789
QY      181 GACCCGATGCGCGCGGGAAGAGCGCGCTTACGCGCGATCTTCTGAGCGACCTACCC 240
DB      7790 GACCCGATGCGCGCGGGAAGAGCGCGCTTACGCGCGATCTTCTGAGCGACCTACCC 7849
QY      241 TCGTTGAGCGCTCTCTTCTTGGCATCTCGCTCGGAGCGCTGCGGATGAGACCTTCA 300
DB      7850 TCGTTGAGCGCTCTCTTCTTGGCATCTCGCTCGGAGCGCTGCGGATGAGACCTTCA 7909
QY      301 CATGACCTCTTCTGAGAGTGTGCTGGAGAGCGCTGAGAGAGCGCGCATGCTCTCATCG 360
DB      7910 CATGACCTCTTCTGAGAGTGTGCTGGAGAGCGCTGAGAGAGCGCGCATGCTCTCATCG 7969
QY      361 GCGCTGTGCGGAGCGAAGCGGAGTGTTCATCGGATCGGCGCGCTCGCATATGAGGCC 420
```


Db 7970 GCGCTGTCGGATCGAAACGGAGATGTTCAATCGGAATCGGCCCTCCGAATATAGACC 8029
Qy 421 GCGCTGCCGAAGCAGACGGCGCTCCGACAGATCGACGCTCATGCGGGCTGGAGACGATG 480
Db 8030 GCGCTGCCGAAGCAGACGGCGCTCCGACAGATCGACGCTCATGCGGGCTGGAGACGATG 8089
Qy 481 CCGACGTCGAGAGCGGGCGGAATCTGATATGCTCCGAGCTGCGAGGGCCGTGTGCGG 540
Db 8090 CCGACGTCGAGAGCGGGCGGAATCTGATATGCTCCGAGCTGCGAGGGCCGTGTGCGG 8149
Qy 541 GTGATATAGGGCTATTTGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 8150 GTGATATAGGGCTATTTGCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8209
Qy 601 TCCGGGAATGCTCCAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 8210 TCCGGGAATGCTCCAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8269
Qy 661 CTCTGTGCTCTCGAAGACCCGGGCGCTGCGCAGAGACGATCGCTGCAAGCATTTTCG 720
Db 8270 CTCTGTGCTCTCGAAGACCCGGGCGCTGCGCAGAGACGATCGCTGCAAGCATTTTCG 8329
Qy 721 GCGAGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 8330 GCGAGGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8389
Qy 781 AGTGAAGCCCGCGCGGAGCGGCGATGAGATTTGGGGGTGATTCGAGGATTCGCGCATTCAT 840
Db 8390 AGTGAAGCCCGCGCGGAGCGGCGATGAGATTTGGGGGTGATTCGAGGATTCGCGCATTCAT 8449
Qy 841 CACGACGCTGCGAGCAGCGGCTGCTGACCGTGCAGAACGAGAGCTCCCAAGAAATCTGCTG 900
Db 8450 CACGACGCTGCGAGCAGCGGCTGCTGACCGTGCAGAACGAGAGCTCCCAAGAAATCTGCTG 8509
Qy 901 AAACGGGCGCTGCGGAGCGCAGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 8510 AAACGGGCGCTGCGGAGCGCAGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8569
Qy 961 GGCACGGGCGAGCGCTTGGTGAACCCATCGAAATCCAAAGCTGAAATGCGGTATAGGCG 1020
Db 8570 GGCACGGGCGAGCGCTTGGTGAACCCATCGAAATCCAAAGCTGAAATGCGGTATAGGCG 8629
Qy 1021 CTGCGGCGAGATGTCGACCGCGCTGCTGATTCGGGTGCTGAGAACCAACCTTGGCAT 1080
Db 8630 CTGCGGCGAGATGTCGACCGCGCTGCTGATTCGGGTGCTGAGAACCAACCTTGGCAT 8689
Qy 1081 CTTGATGATTCGCTGCGGAGATCACTGGGCTGCTGAAGTCTGCTTCTCCCTTCAGACGGG 1140
Db 8690 CTTGATGATTCGCTGCGGAGATCACTGGGCTGCTGAAGTCTGCTTCTCCCTTCAGACGGG 8749
Qy 1141 CAGATTCCTGCGACCTTCAACCGCGAGCGCTGAACCCCGGATTCATATGGGTATCTT 1200
Db 8750 CAGATTCCTGCGACCTTCAACCGCGAGCGCTGAACCCCGGATTCATATGGGTATCTT 8809
Qy 1201 CCGCTGACCTGTCACGCGCGCGCGACACGCTGCGCGGACTGGATATACCGCGGACGGGCG 1260
Db 8810 CCGCTGACCTGTCACGCGCGCGCGACACGCTGCGCGGACTGGATATACCGCGGACGGGCG 8869
Qy 1261 GGGGTGAGCTGCTTGGCATGAGCGGGAACAAACGCTGACGCTGCTGCTGCTGCTGCTGCTG 1320
Db 8870 GGGGTGAGCTGCTTGGCATGAGCGGGAACAAACGCTGACGCTGCTGCTGCTGCTGCTGCTG 8929
Qy 1321 GCGGAGAGTGCACACCGCGCGCGCGGAGCGGACCGGAGAGCTGCTGCTGCTGCTGCTGCTG 1380
Db 8930 GCGGAGAGTGCACACCGCGCGCGCGGAGCGGACCGGAGAGCTGCTGCTGCTGCTGCTGCTG 8989
Qy 1381 AGGACCGGCTACGCTTGGATGACAGGCGCGCGCGCTGCGGACCATCTGAGACCTAC 1440
Db 8990 AGGACCGGCTACGCTTGGATGACAGGCGCGCGCGCTGCGGACCATCTGAGACCTAC 9049
Qy 1441 CTTTGCAGTGTCTGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500

Db 9050 CTTTGCAGTGTCTGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9109
Qy 1501 CACCGGCTCGGCTGCGGCGGACGATCGAGAGGAGGAGCTGCGGAGACCTTGAACGCTGCG 1560
Db 9110 CACCGGCTCGGCTGCGGCGGACGATCGAGAGGAGGAGCTGCGGAGACCTTGAACGCTGCG 9169
Qy 1561 GCGGAGGAGACGATGCGCGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1620
Db 9170 GCGGAGGAGACGATGCGCGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9229
Qy 1621 CTTGCTTTCTTTCACCGGACAGGGGCGGAGACGCTGCGGACATGAGGCGGCTGCTGCTGCTG 1680
Db 9230 CTTGCTTTCTTTCACCGGACAGGGGCGGAGACGCTGCGGACATGAGGCGGCTGCTGCTGCTG 9289
Qy 1681 GATGATATGCTCGGCTTCCGAGAGGCTTCAACCTGTCGCTGAGAGGCTTCAACAGAGAG 1740
Db 9290 GATGATATGCTCGGCTTCCGAGAGGCTTCAACCTGTCGCTGAGAGGCTTCAACAGAGAG 9349
Qy 1741 CTGACCGGCGGCTCCGAGAGTATGAGGCGGACCGGCGGAGCGGCGGCTG 1800
Db 9350 CTGACCGGCGGCTCCGAGAGTATGAGGCGGACCGGCGGAGCGGCGGCTG 9409
Qy 1801 CTGACCGAGACAGCTTCAACCGGCGGCTGCTTCAACCTTCAATATGCGCTGCGGCGG 1860
Db 9410 CTGACCGAGACAGCTTCAACCGGCGGCTGCTTCAACCTTCAATATGCGCTGCGGCGG 9469
Qy 1861 CTGTCGCGGCTGCTGAGGCTGAGCGGAGGCTGCTGCGGCGGAGCTGAGCATGCTGAGCTG 1920
Db 9470 CTGTCGCGGCTGCTGAGGCTGAGCGGAGGCTGCTGCGGCGGAGCTGAGCATGCTGAGCTG 9529
Qy 1921 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Db 9530 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9589
Qy 1981 CCGCGGCGGCTGAGTACAGGCGCTGCGGCGGCGGCGGAGTGTGCTGCTGAGGCGGCGG 2040
Db 9590 CCGCGGCGGCTGAGTACAGGCGCTGCGGCGGCGGCGGAGTGTGCTGCTGAGGCGGCGG 9649
Qy 2041 GAGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
Db 9650 GAGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9709
Qy 2101 AACGCTCGGACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
Db 9710 AACGCTCGGACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9769
Qy 2161 GCGATGCGCGGCGGCGGCGGCGGAGCAACCAAGGCGCTTCAACGCTGCTGCTGCTGCTGCTG 2220
Db 9770 GCGATGCGCGGCGGCGGCGGCGGAGCAACCAAGGCGCTTCAACGCTGCTGCTGCTGCTGCTG 9829
Qy 2221 CCGCTCATGCGCGGAGTGTGAGAGCGGCTGCGGCGGCTGCGGAGTGTGAGGCTACCGG 2280
Db 9830 CCGCTCATGCGCGGAGTGTGAGAGCGGCTGCGGCGGCTGCGGAGTGTGAGGCTACCGG 9889
Qy 2281 CCGGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
Db 9890 CCGGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9949
Qy 2341 TCGCGGCGGCTATTTGGGTGCGGACCGCGGAGAGTGTGCTGCTGCGGAGTGTGAGG 2400
Db 9950 TCGCGGCGGCTATTTGGGTGCGGACCGCGGAGAGTGTGCTGCTGCGGAGTGTGAGG 10009
Qy 2401 GCGCTGACCGGCGGCGGCGGAGCACTTGTGAGAGGCTGCTGAGAGTGTGAGAGTGTGCTGCTG 2460
Db 10010 GCGCTGACCGGCGGCGGCGGAGCACTTGTGAGAGTGTGCTGAGAGTGTGAGAGTGTGCTGCTG 10069
Qy 2461 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
Db 10070 GGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10129
Qy 2521 GCGGCTGACGAGCGGCGGAGCGCTGCTGAGGCGCTGCGGCGGCTGCTGCGGCTGCTGCTGCTG 2580
Db 10130 GCGGCTGACGAGCGGCGGAGCGCTGCTGAGGCGCTGCGGCGGCTGCTGCGGCTGCTGCTGCTG 10189

[illegible]

Oy		3661	CCGGGCGGGGGTCCGTGCCTCTTTGAAGACCTCTGGACTCTATGCCGACGCGTGTC	3720
Dd		11270	CCGGCGGGGGTCCGTGCCTCTTTGAAGACCTCTGGACTCTATGCCGACGCGTGTC	11322
Oy		3721	ACCCTCCCCCATCGCAGCGCTCCCGATGACCGGTGTGCGCCGANTCGTTCCGGAGCATG	3780
Dd		11330	ACCCTCCCCCATCGCAGCGCTCCCGATGACCGGTGTGCGCCGANTCGTTCCGGAGCATG	11389
Oy		3781	GCGCAGGCGCAGCATCTTGGGAAAGCTGTACTCAAGCTGGGTGACCCGGAGGTCCAGATC	3840
Dd		11390	GCGCAGGCGCAGCATCTTGGGAAAGCTGTACTCAAGCTGGGTGACCCGGAGGTCCAGATC	11449
Oy		3841	CGATTTCACACCCACGAGGCGCGCGCCGCTGCCACGCGGGGATGGGAGACCTGCACAGG	3900
Dd		11450	CGATTTCACACCCACGAGGCGCGCGCCGCTGCCACGCGGGGATGGGAGACCTGCACAGG	11509
Oy		3901	CTCCAGCTCACGTGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCCTCGTAGCGCAG	3960
Dd		11510	CTCCAGCTCACGTGCGCGCGCGCGCGCGCGCGCGCTGAGAGCGTTCCTCGTAGCGCAG	11569
Oy		3961	GTCCTGCAGAGTGTGCGCACGCGCCGAATCAAGGTGCGCGCGGAGCGCTGTACCCGC	4020
Dd		11570	GTCCTGCAGAGTGTGCGCACGCGCCGAATCAAGGTGCGCGCGGAGCGCTGTACCCGC	11629
Oy		4021	CTCGGCAATGCAATCCGCTCATATGCGCGGTGAGACCTGCGAATCTGATGAGGCGAGCCTCAAG	4080
Dd		11630	CTCGGCAATGCAATCCGCTCATATGCGCGGTGAGACCTGCGAATCTGATGAGGCGAGCCTCAAG	11689
Oy		4081	CTGAAAGCTGTGCAGACGAGCTTCTGTCCAAGCTCCCGCAATATGCGCTGTGTGGCCCAAAC	4140
Dd		11690	CTGAAAGCTGTGCAGACGAGCTTCTGTCCAAGCTCCCGCAATATGCGCTGTGTGGCCCAAAC	11749
Oy		4141	CTGTTGAGTCTCTGCGCACAGCTCTTCTCTTGAGACGGGTGCGCGGAGAACCTACCG	4200
Dd		11750	CTGTTGAGTCTCTGCGCACAGCTCTTCTCTTGAGACGGGTGCGCGGAGAACCTACCG	11809
Oy		4201	GCAGGCGCTGCAAAACGACTTCGTCTCATATGCGGGCGAGATCAAGACTGGGAAATCATTTGCC	4260
Dd		11810	GCAGGCGCTGCAAAACGACTTCGTCTCATATGCGGGCGAGATCAAGACTGGGAAATCATTTGCC	11869
Oy		4261	CTATGGA 4266 	
Dd		11870	CTATGGA 11875	
RESULT 2				
ID	AAA29349			
AAA29349 standard; DNA; 71989 BP.				
XX	AAA29349;			
AC				
XX				
DT	12-SBP-2000 (first entry)			
XX				
DE	Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.			
XX				
KM	Epothilone; polyketide synthase; epoA; epob; epoc; epod; epof; epog; epok; epok; P450 epoxidase; ORF A; ORFB; promoter; enhancer; anti-fungal; tubulin polymerization assay; anti-tumor; cytostatic; ds.			
XX				
OS	Sorangium cellulosum.			
XX				
FH	Key . Location/Qualifiers			
FT	CDS 3..992			
FT	/cag= a			
FT	/label= ORF A			
FT	/product= transposase			
FT	/note= "not part of the PKS"			
FT	989..1501			
FT	/cag= b			
FT	/label= ORF B			
FT	/product= transposase			
FT	/note= "not part of the PKS"			
FT	1998..6263			
CD				
FT				

```
FT      /*tag= c
FT      /label= epoc_gene
FT      /note= "encodes the loading domain"
FT      2031..3548
FT      /*tag= d
FT      /note= "encodes ketide synthase (KS-Y) of the loading
FT      domain"
FT      3621..4661
FT      /*tag= e
FT      /note= "encodes acyl transferase (AT) of the loading
FT      domain"
FT      4917..5810
FT      /*tag= f
FT      /note= "encodes enoyl reductase (ER) of the loading
FT      domain, potentially involved in formation of the
FT      thiazole moiety"
FT      5856..6155
FT      /*tag= g
FT      /note= "encodes acyl carrier protein (ACP) of the loading
FT      domain"
FT      6260..10493
FT      /*tag= h
FT      /label= epob_gene
FT      /note= "encodes module 1, the NRPS module"
FT      2031..3548
FT      /*tag= i
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      2031..3548
FT      /*tag= j
FT      /note= "encodes condensation domain C2 of the NRPS
FT      module"
FT      6861..6887
FT      /*tag= k
FT      /note= "encodes heterocyclization signature sequence"
FT      6861..6887
FT      /*tag= l
FT      /note= "encodes condensation domain C4 of the NRPS
FT      module"
FT      7358..7366
FT      /*tag= m
FT      /note= "encodes condensation domain C7 (partial) of the
FT      NRPS module"
FT      7898..7921
FT      /*tag= n
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      7898..7921
FT      /*tag= o
FT      /note= "encodes adenylation domain A1 of the NRPS module"
FT      8261..8308
FT      /*tag= p
FT      /note= "encodes adenylation domain A3 of the NRPS module"
FT      8411..8422
FT      /*tag= q
FT      /note= "encodes adenylation domain A4 of the NRPS module"
FT      8861..8905
FT      /*tag= r
FT      /note= "encodes adenylation domain A6 of the NRPS module"
FT      8986..8983
FT      /*tag= s
FT      /note= "encodes adenylation domain A7 of the NRPS module"
FT      9090..9179
FT      /*tag= t
FT      /note= "encodes adenylation domain A8 of the NRPS module"
FT      9183..9992
FT      /*tag= u
FT      /note= "encodes oxidation region for forming thiazole"
FT      10121..10138
FT      /*tag= v
FT      /note= "encodes adenylation domain A10 of the NRPS
FT      module"
FT      10261..10306
FT      /*tag= w
FT      misc_RNA

FT      /*tag= x
FT      /label= epoc_gene
FT      /note= "encodes module 2"
FT      10654..12033
FT      /*tag= y
FT      /note= "encodes KS2, the KS domain of module 2"
FT      12250..13287
FT      /*tag= z
FT      /note= "encodes AT2, the AT domain of module 2"
FT      13327..13899
FT      /*tag= aa
FT      /note= "encodes dehydratase (DH) 2, the DH domain of
FT      module 2"
FT      14962..15756
FT      /*tag= ab
FT      /note= "encodes ketoreductase (KR) 2, the KR domain of
FT      module 2"
FT      15763..16008
FT      /*tag= ac
FT      /note= "encodes ACP2, the ACP domain of module 2"
FT      16134..37907
FT      /*tag= ad
FT      /label= epod_gene
FT      /note= "encodes modules '3-6"
FT      16425..17606
FT      /*tag= ae
FT      /note= "encodes KS3"
FT      17817..18857
FT      /*tag= af
FT      /note= "encodes AT3"
FT      19581..20396
FT      /*tag= ag
FT      /note= "encodes KR3"
FT      20424..20642
FT      /*tag= ah
FT      /note= "encodes ACP3"
FT      20706..22082
FT      /*tag= ai
FT      /note= "encodes KS4"
FT      22296..23336
FT      /*tag= aj
FT      /note= "encodes AT4"
FT      24069..24647
FT      /*tag= ak
FT      /note= "encodes KR4"
FT      24867..25151
FT      /*tag= al
FT      /note= "encodes ACP4"
FT      25203..26576
FT      /*tag= am
FT      /note= "encodes KS5"
FT      26793..27883
FT      /*tag= an
FT      /note= "encodes AT5"
FT      27966..28574
FT      /*tag= ao
FT      /note= "encodes DHS"
FT      29433..30287
FT      /*tag= ap
FT      /note= "encodes ER5"
FT      30321..30869
FT      /*tag= aq
FT      /note= "encodes KR5"
FT      31077..31373
FT      /*tag= ar
FT      /note= "encodes ACP5"
FT      31440..32807
FT      /*tag= as
FT      /note= "encodes KS6"
FT      33018..34067
FT      misc_RNA
```


Db 3678 GATGATGCGCCGCTTCCGAGGCGTTTCAGCTGTGCGTGAAGCTTTCAACAGAG 3737
Qy 1741 CTCGACCGGCGGCTCCGAGAGTGTGTGGCCGAACCGGACGAGTGAAGCGCGCTG 1800
Db 3738 CTCGACCGGCGGCTCCGAGAGTGTGTGGCCGAACCGGACGAGTGAAGCGCGCTG 3797
Qy 1801 CTCGACGAGACGCTTCAACCGGCGGCTGTTCACCTTGAATATGCGTCCGCG 1860
Db 3798 CTCGACGAGACGCTTACCGAGCGGCGCTGTTCACCTTGAATATGCGTCCGCG 3857
Qy 1861 CTGTGCGGCTGTGTGGAGTGTGAAGCGGAGTGTGTGGCCGAGTGAAGTGTG 1920
Db 3858 CTGTGCGGCTGTGTGGAGTGTGAAGCGGAGTGTGTGGCCGAGTGAAGTGTG 3917
Qy 1921 GTGCGTGTGCTGTGTGGAGGCGTGTTCGCTTGAAGAGCGGAGTGTCTGTGGCTG 1980
Db 3918 GTGCGTGTGCTGTGTGGAGGCGTGTTCGCTTGAAGAGCGGAGTGTCTGTGGCTG 3977
Qy 1981 GCGGCGGCGCTGTGTGAAGCGGCTGTGGCGGCGGAGTGTGTGAAGCGGCG 2040
Db 3978 GCGGCGGCGCTGTGTGAAGCGGCTGTGGCGGCGGAGTGTGTGAAGCGGCG 4037
Qy 2041 GAGGCGGAGT 2100
Db 4038 GAGGCGGAGT 4097
Qy 2101 AACGCTCGGAGCGAGT 2160
Db 4098 AACGCTCGGAGCGAGT 4157
Qy 2161 GCGAGT 2220
Db 4158 GCGAGT 4217
Qy 2221 CCGCTCATGCGCGGCTGT 2280
Db 4218 CCGCTCATGCGCGGCTGT 4277
Qy 2281 CCGGCGGAGT 2340
Db 4278 CCGGCGGAGT 4337
Qy 2341 TCGCGGAGT 2400
Db 4338 TCGCGGAGT 4397
Qy 2401 GCGCTGACGCGGCGGCTGT 2460
Db 4398 GCGCTGACGCGGCGGCTGT 4457
Qy 2461 GCGCTGAGT 2520
Db 4458 GCGCTGAGT 4517
Qy 2521 GAGGCTGACGAGCGGCGGCTGT 2580
Db 4518 GAGGCTGACGAGCGGCGGCTGT 4577
Qy 2581 CTGT 2640
Db 4578 CTGT 4637
Qy 2641 CTTTGGAGCGGCGGCTGT 2700
Db 4638 CTTTGGAGCGGCGGCTGT 4697
Qy 2701 GCGCGT 2760
Db 4698 GCGCGT 4757
Qy 2761 GACGCGGAGGCTGT 2820

Db 4758 GACGCGGAGGCTGT 4817
Qy 2821 GAGGCGGCGGCGGCTGT 2880
Db 4818 GAGGCGGCGGCGGCTGT 4877
Qy 2881 CTGT 2940
Db 4878 CTGT 4937
Qy 2941 GACGCGGCGGCGGCTGT 3000
Db 4938 GACGCGGCGGCGGCTGT 4997
Qy 3001 CTGT 3060
Db 4998 CTGT 5057
Qy 3061 GCGT 3120
Db 5058 GCGT 5117
Qy 3121 GAGGCTGT 3180
Db 5118 GAGGCTGT 5177
Qy 3181 CTGT 3240
Db 5178 CTGT 5237
Qy 3241 GACGAGT 3300
Db 5238 GACGAGT 5297
Qy 3301 GT 3360
Db 5298 GT 5357
Qy 3361 GCGAGCGGCGGAGGCTGT 3420
Db 5358 GCGAGCGGCGGAGGCTGT 5417
Qy 3421 GCGT 3480
Db 5418 GCGT 5477
Qy 3481 GT 3540
Db 5478 GT 5537
Qy 3541 GCGGCGGTTGT 3600
Db 5538 GCGGCGGTTGT 5597
Qy 3601 CCGT 3660
Db 5598 CCGT 5657
Qy 3661 CCGGCGGAGT 3720
Db 5658 CCGGCGGAGT 5717
Qy 3721 ACCGCTGT 3780
Db 5718 ACCGCTGT 5777
Qy 3781 GCGGAGCGGAGGCTGT 3840
Db 5778 GCGGAGCGGAGGCTGT 5837
Qy 3841 GGTATTTCCACCGAGGCTGT 3900
Db 5838 GGTATTTCCACCGAGGCTGT 5897

Db 43974 GGGGGTCTTCGCGCCCTTTCACGAGGAGCTACCTCTGTTGTC-----TCTCGGA 44027
Qy 441 GTCCGAGAGATCGACGCTCAGGCGGGGAGCATGCGGAGCGTGGAGCGGGCG 500
Db 4428 TCTCAAGCCATCAACGCTCTTACCGAGCAGGCAAGGCAACACGCGGCTGGACG 44087
Qy 501 AATCTGATATCCCTCGGCTGCGAGGCGCGTGTGTCCGGTGAATACGGCTTATCTC 560
Db 4408 ATTGCTCATCTGCTGAGCCTGCAAGGGCCAGCTTTGTGTGAGCAGCGGCTGCTCTC 44147
Qy 561 CTGCTGTGCGCTTCACTGCGCTCTCAAGCTTGGCTTCGGGATGCTCCAGCGC 620
Db 44148 CTCCTGTCTCCGCTTCACTGCGGCTGTCAAGGCTTCTGAATCGTGAATCAGATGCG 44207
Qy 621 CTTGCTGTGCGGATGCTGATGTTGTGCGCGAGCAGCCCTGCTGCTCGAAGAC 680
Db 44208 GGTGCTGAGCGGCTGACGTTGTTGCTACGCTACGACCTCTGCTATGCGGCTGAAGAT 44267
Qy 681 CCGGCGCTGCGCAGGAGCGGTCTGTCAGGCAATTTTCGCGAGGCGAGTGGGTTCCG 740
Db 44268 GCACTCATGCGCGCGAGCGGCGCTGCAAGCTTGAAGCGCGCGCGATGGATCGC 44327
Qy 741 ACGAGCGAAGGAGGCGCGCTGCTGCTCTCAAGCGCTCAGTGAAGCGCGCGAGCG 800
Db 44328 CATGGGTGAGGCTGTGCGCTGTGATCTCAAGCGCTTGTGAGTGCAGCGCGATGG 44387
Qy 801 CGATCGATATTGCGGATGATTCAGGATCCGCGATCAATCAGACGCGTGGAGAGCGG 860
Db 44388 TGAATCGATCTTGGGAGTGAATCAAGCTTCGCGATCAATCAAGCGCGCGCAATGG 44447
Qy 861 TCTGACCGTGCAGACGAGAGCTCCAGAAATCGTGTGAACGCGGCTTGCGGAGCGC 920
Db 44448 GCTACACGCGCGAGTGGCGCTTCCAGAGGTGTTGCGTGAAGGCGCTGAGCGCGCGC 44507
Qy 921 AGGCTGCGCGCTTCTGCTGCTGATATGTCAGGAGCAGCGACGCGGAGCGCTTGG 980
Db 44508 TGGCATCTCGCGCTGCGAGGTAGCTAGTGAAGCCAGCGGTTCGCGGAGCAGCGCTGG 44567
Qy 981 TGACCCCATGGAATCCAAAGCTGGAATGCGGTATACCGGCTCGGCGAGATGTCGACAC 1040
Db 44568 CGATTCGATGAGGTGAGGCGCTTCGCGAGGTCAATCGGAGCGACCACTTCGCGGCC 44627
Qy 1041 GCGCTGCTGATCGAGTCTGATGAAGACCAACCTTGGCATCTGATGATGCTGCGGAGAT 1100
Db 44628 TCCGCTGCTCTGAGTTCCGTAAGGCCAATCTGGGACATACCGCGCGGCGGCGGGGT 44687
Qy 1101 CACTGGGCTGCTGAAGTCTGCTTGTCTTCAGCAGCGGAGATTCCTGCGCACTTCCA 1160
Db 44688 CGCGGGGCTGATCAAAAGTCTGCTCTCATGACAGACGAGCGGATTCCTCTCAACTGCA 44747
Qy 1161 CGCGCAGGCGCTGAACCCCGGATCTCAATGGGATGATCTTGGCGGACCGTCAAGCGCGC 1220
Db 44748 CTACCGGAGCTCAACTCAACATCTCTTGAAGAACCTTCTGATCTTCGAGAGA 44807
Qy 1221 CCGGACACCTGCGCGAGCTGAATACCGCGCGACGCGGCGGAGTGAAGCTTCTGCGCAT 1280
Db 44808 GCTGATCTCTGCGCTGCGGAGAGGGGCGCGGATTCGCAAGGATCAAGTCTCTGCTG 44867
Qy 1281 GAGCGGAGCAACGCGCAGTGTGTGTAAGAGGCGCGCGCGGAGCTGACACCGCGC 1340
Db 44868 GTCAAGGAGCAACGCGCAGTGTGTGTAAGAGGCGCGCTGCGCGCCCAAGAGCTCGCGC 44927
Qy 1341 GCGCGCGAGGAGCCGCGAGAGCTGTGCTGTGCGGAGAGCCGCTGACGCGTGGGA 1400
Db 44928 GCGCGTTGAGGCGTCCGAGTCACTCTGACCTGTGCGCTCGAGCCGAGCGCGCTGCG 44987
Qy 1401 TGCAAGGCGCGCGGCTGCGGAGCATCTGAAGACCTTACAGTGTCTTGAGCGGA 1460
Db 44988 CGAGTTGCGCGAGGCTGTAAGACATATGCTCTGCGCGAGCGGAGATCTGCCCATGCGCGA 45047
Qy 1461 TGTGCGTTCACTGTGCGAGAGCGCGAGCGAGTGAAGACCGGCTGCGGCTGCGCGCGC 1520

Db 45048 TCTCTGCTACTCGGCAACACGCGGAGTGTGAAATTGAGACCATCTGCGGCTCCCCG 45107
Qy 1521 GACGTGAGGAGAGGGGCTGCGGAGCGCTTGAAGCTGCGCGGAGGAGACAGAGTGC 1580
Db 45108 GGGGACAGCGGCCCAAGTTGAGCGAGCGTCTCCGCTTTCTTGAAGGAGACGCGCAAGC 45167
Qy 1581 CGGTGCGGTGTGAGTATCGCCGATTCCTCAAGCGGCAAG--CTGCGCTTTCTTTAC 1637
Db 45168 AGGACGCTGTAAGGACAGGTCACGAGCGCGCGTGGGCGGAGATCGGCTCTGTTAC 45227
Qy 1638 CGACAGAGGAGGAGAGCGCTGAGCATGGGCGCTGAGGCTGATGATGATGCTCGGCT 1697
Db 45228 GGGGACAGGCTGCGAGTACCGGAGCATGGGCGGCACTTACAGACAGCCCACTT 45287
Qy 1698 CCGGAGGCGCTTGAACCTGTGAGGAGCTGTTCAACAGAGGCTGACCGGCGCTCG 1757
Db 45288 CCGGCGCTCTCATGATGATGAGTGAAGTGTCTGCGCCCTTGTCTGCGGCGCTCATCT 45347
Qy 1758 CGAGTATGTGAGCGCAACCGGCGAGCTGACGCGCGCTGCTCAACAGACGCTT 1817
Db 45348 GTGCTGATCTACCCGAGCG-----GGGCAAGCTTCTCGATCAAGAGCGGCTT 45401
Qy 1818 CACCCAGCGCGCTGTTCACTTCAATATGCTGCGCGCGCTGCTGCGGCTGCTGCGG 1877
Db 45402 CACGACCGCGCTGTTGCTGCTGAGTGGGCTGCGGAGCTTGGCGCTGCTGCGG 45461
Qy 1878 TGTAGACCGGAGTGTGCTGCGCGCATAAGCATGATGATGCTGCTGCTGCTGCTG 1937
Db 45462 GCTGTACACCGTGTGATGAGGCTCACAGCTGATGATGATGATGATGATGATGATG 45521
Qy 1938 GGGGCTGTTCTCGTTAGAGACGCGGCTTCTGAGGCTGCGCGCGCGCTGATGCA 1997
Db 45522 TGGGCTGTTCACTTCAGAAAGGCTTGGCTCATGTCGAGAGCGCGCGCTGATGCA 45581
Qy 1998 GCGCTGCGCGCGCGCGCGGATGATGATGATGATGATGATGATGATGATGATG 2057
Db 45582 GTGCTGCGCGCGCGCGCGGAGTGGCGCGCTTTCACGAGCGGCGCTCGTACCGCA 45641
Qy 2058 TGGGCTGCGCGCGCGCGGAGTGGCGCGCTTTCACGAGCGGCGCTCGTACCGCA 2117
Db 45642 GCGGATGCTCTTACGATGAGAGTCTCATGTCGCGCTTCAATGATCTTCCGAGAC 45701
Qy 2118 GGTGATGCGCGCGCGCGGCAACCGTGAATGCGATGCGCGCGGAGTGGCGCGCG 2177
Db 45702 GGTGATCTTGGCGATGAGCAAGCGTGAAGCGATTCGCGGCGCTTGTCCGCGAGG 45761
Qy 2178 GCGCGAACAAGCGCTTCCAGCTCTGCGATGCTTCACTCAACGCTCAATGCGCGGAT 2237
Db 45762 CGTAAAGTTCGCGCGCTGAGTGTCTCCAGCGCTTCACTTCCGCTCATGAGACCAT 45821
Qy 2238 GCTGAGGCGCTTGGGCGTGTGCGAGTGGAGTCAACCGGCGCGCGCTGATGCTCT 2297
Db 45822 GCTGAGTGTCTTCAAGCGCGCGCGCGAGAGACGCGCTTCAATGCGCGCGAGATCA 45881
Qy 2298 GGTGAGCATCTGAAGCGGAGGCTTGAAGAGTGAAGTCAAGCGGCTGATGCTCT 2357
Db 45882 GATCTGAACCTTCAACCG-----TGGGCGCTGAGAGAGTTCGCGCGCGGCTTGG 45935
Qy 2358 GCGCGAAGCGGAGAGTGTGCGCTTTCGAGATGAGTGAAGCGCTGCAACGCGCGG 2417
Db 45936 TCGACATGCGCGGTGAACCGGCTTCTCGAGTGAAGTGGCGCGCTCAAGAGCAGTGG 45995
Qy 2418 TGGCGGACCTTGTGAGGCTGAGTTCGGAATGAGACGCTGCGCGCTGAGTGTGCTG 2477
Db 45996 GCTTCTGCTGCGGAGGATGAGTCCGGAATGCACTTGTCTGCGCATCGAAGCGG 46055
Qy 2478 CATGCGGAGCGCGCGCGCGCTGCTGCAATGCTGCGCGCTGAGGCGTGAAGAGCGG 2537
Db 46056 TCTGCGGAGAGGCTGAGTGTGCTTGTGCTTCCCTTCAAGAGCAAGAGCGAGTGG 46115
Qy 2538 GACCTGCTGAGGCGCTGCGCGGCTGCTGAGGCTGAGGCTGAGGCTGCTTCTGCGCG 2597
Db 46116 AGTCTCTGCAACGTTGAGTGGCTCTGTGCGAGAGGCTTCCGTTCACTGGGAGG 46175

OY	2598	---CCTCTTCCCTCAGAGGGGGCGCGGGGAGGCCCTGACCGTACCTTGGAGCGCA	2654
Db	46176	GTTGACAGACGACACTACCGCGCTCGGCGGGTCTCCCTCCCAATTATTCGTTAGACACA	46233
OY	2655	GCGCTACTGATCGACACGAAAGCCGACGACGCGCGCGTGGCGACCGCCGTCTCGGG	2714
Db	46236	GCGGTTCGCGCTC-----AAGCAGATGTAGTGAACGGGGGTGTGGCGCCGACGGCGTCAAG	46288
OY	2715	AGCGGTCACACGAGGTCGACAGAGGGGGCGCGGTCCGGCGGGCGACCGGGCGACCGC	2774
Db	46289	TCGTGGGTGATCCCTCTGTGGGATCGGCGCTCGCGCGTGGACGATGAGACGACACTGTGT	46348
OY	2775	TCGGCTCGACCATCCG	2790
Db	46349	TCGAATCGAAGTCCG	46364
RESULT 4			
AACS5785			
ID	AACS5785	standard; DNA; 3978 BP.	
XX			
AC	AACS5785;		
XX			
DT	19-JAN-2001	(first entry)	
XX			
DE	Type I polyketide synthase orf8.		
XX			
KW	Mitomycin, biosynthesis; mitosome ring system; antibiotic; anti-cancer;		
KW	anti-inflammatory; immune-enhancer; immunosuppressant; asthma;		
KW	chronic obstructive pulmonary disease; respiratory inflammation;		
KW	fungicide; pesticide; ds.		
XX			
OS	Streptomyces lavendulae.		
XX			
PN	WO200053737-A2.		
PD	14-SEP-2000.		
XX			
PF	10-MAR-2000; 2000WO-US06394.		
XX			
PR	12-MAR-1999; 99US-0266965.		
XX			
PA	(MINU) UNIV MINNESOTA.		
PA	(SHER/) SHERMAN D H.		
PA	(MAOY/) MAO Y.		
PA	(VARO/) VAROGU M.		
PA	(HEM/) HE M.		
PA	(SHEL/) SHELTON P C.		
XX			
PI	Shelman DH, Mao Y, Varoglu M, He M, Sheldon PC;		
XX			
DR	WPI, 2000-601980/57.		
XX			
PT	Novel nucleic acid molecule comprising mitomycin biosynthetic gene		
PT	cluster useful for cloning mitomycin biosynthetic genes for elucidating		
PT	the molecular basis of mitosome ring system biosynthesis		
XX			
PS	Example 1; Page 251-253; 39pp; English.		
XX			
CC	This invention relates to isolated and purified nucleic acid molecules		
CC	from the mitomycin biosynthetic gene cluster. Mitomycins are a group of		
CC	natural products that contain a variety of functional groups, including		
CC	amino benzoxinone and axiridine ring systems. The S. lavendulae		
CC	mitomycin biosynthetic gene cluster comprises 47 mitomycin genes		
CC	spanning 55kb of DNA. The invention includes an expression cassette		
CC	comprising a mitomycin biosynthetic gene operably linked to a promoter,		
CC	and host cells transformed with the cassette. The nucleotide, and protein		
CC	sequences and the transformed host cells of the invention result in		
CC	antiaesthetic, antiinflammatory, cytostatic, immunomodulatory, and		
CC	antibiotic activities. The nucleotide sequences are used to elucidate the		
CC	molecular basis for the biosynthesis of the mitosome ring system, as well		
CC	as to engineer the biosynthesis of novel natural products, e.g.		

	Query Match	19.1%	Score 81.4.2;	DB 121;	Length 3978;
	Best Local Similarity	57.6%	Pred. N1.1e-121;		
	Matches 1549;	Conservative	0;	Mismatches 1108;	Indels 30; Gaps 4
CC	antibiotics, anti-inflammatory agents, anti-cancer agents,				
CC	immune-enhancers, immunosuppressants, agents to treat asthma, chronic				
CC	obstructive pulmonary disease as well as other disease involving				
CC	respiratory inflammation, or cholesterol-lowering agents or as crop				
CC	protection agents (e.g. fungicides or insecticides) as well as				
CC	biopolymers, e.g., in packaging or biomedical applications, or to engineer				
CC	PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and				
CC	AAB32485-332542 represent mycomycin biosynthetic gene cluster DNA				
CC	sequences and encoded proteins. Sequences AAC55812-C55814,				
CC	AAC55850-C55866 and AAC55862-C55869 represent PCR primers used in the				
CC	cloning of the mycomycin biosynthetic genes.				
XX	Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 other;				
XX					
QY	18 CGAGCGCGACGCCGGAAGATCCGATTGCGATGTCGAGACGACTTGCCCTGCCGGTGC 77				
DB	90 CGAGGACCCCGGCGCCCGAGCCCGTCGCATCGTGGGCAATGGCGTCCGCTTCCCGGGGA 149				
QY	78 CGTATCGATCGAGCGGGTTCTGAGCGCTCCTCGAGGAGCTCGCGCGACACCGTCGGGG 137				
DB	150 CTTGCGCATCGCCGACGACACTGTGCACTGTGTCGCCGAGGCCGGAGCGCCTTACCGA 209				
QY	138 AGTCCCGCCGGAAC--GCTGGAGATGACAGACGCGTGTGATTATCCGACCCCGATGCC 194				
DB	210 GTTCCCGCCGACCGGGGCTGGAGAGTGCAGCGCGCTTACAGACCCCGAGCCGGACCCC 269				
QY	195 GGGAGAGACGCCCTTATGACCGCGCATCTTTCCTGAGCGACGTAGCGTCTTGAAGCCGC 254				
DB	270 GGGCAGAGACGTACCGCGCGCCACAGCGCGCTTCTTCAAGACGCGCGGATTTGACGCGCG 329				
QY	255 CTTCTTCCGACACTTCGCTCGCGAAGCCCTGCGATGAGACCTCGACATCGACTTTGCT 314				
DB	330 CTTCTTCCGACATACGCGCGCGAGCGGCTCGCATGAGACCGGACGAGCGCGCATGATCAT 389				
QY	315 GGAGGTGTGCTGGGAGGCGCTGAGAGAACGCCGCGATCGCTCCATGGGCGCTGTCGGTAC 374				
DB	390 GGAGGTCTCTCGGGAGGCGTTTCGACAGACGGCGGCTTCGACGGACACCTTCGGGGCGA 449				
QY	375 GGAACCGGAGAGTGTTCATCGGAGATGGGCCCGTCGATATGAGCGCGCTCCCGCAAGC 434				
DB	450 GGACGTCCGCGCTCTTCGTGCGGTCCAAAGACAGACACTACGTATCAAGTCTCGAGCGC 509				
QY	435 GACGCGCTCCGACGAGATGACGCTCATAGCGGGCTGGGAGACGATCCGACGCTGGAGC 494				
DB	510 GCGGAGCATCGCGCGAGGGCTTC-----ATCGGGAGCCGGAACAATCCGCGACGATCCTTC 563				
QY	495 GGGCGCGAATCTCGATATGCGCTCGGCGTCGAGGGCGCGTGTGCGGGTGAATACGGCGTA 554				
DB	564 CGGCGCGGTGCTACACCTTTCGCTTTCGAGGGGCCCGCGCGTGTCCGTGACACCGCTGC 623				
QY	555 TTTCGTCTCGCTGTGAGCGCTTCATCTTGAGCTTCGACGCTTTCGCGCTCGGGAAATGCTC 614				
DB	624 CTCTCTCTCGCTGTGCGCTGACACTTGACCGCGACGATCCGCGGAGGGGAGTGTCTC 683				
QY	615 CACGCGCCTTGCTGGTGGGGGTATGCTGTATGTGTTCGCGAGACACCTCGTGTGCTCTC 674				
DB	684 CTTGCGCTTGGCGGGCGGGCGGACGCGTATGGCCAGCGCGGACCGCTTTCATGAGTTTCAG 743				
QY	675 GAAGACCGGGGCGCTGCGCAGGAGACGCTGCGTCGCAAGACATTTTTCGCGGAGGCGGATGG 734				
DB	744 CGGCGAGGGGGCTTGGGCTCCCGAGCGCGCTGCGAAGTCTTCTCGGCGACCGCGGACGG 803				
QY	735 GTTTCGACGAGCGGAGGGGTGCGCGCTGTGTCTCTCAAGCGGCTTCAGTGTAGCCCGCGC 794				
DB	804 CACCACTGTGTCCGAGGGCGCGGCGGTCTGCTGTGCGCGCGCTTCGCGAGCGCCGCGC 863				
QY	795 GGAACGCGATCGGAATATGGCGGTGATTCGAGGATCCGCGATCAATCAAGAGTGTGCGAG 854				
DB	864 CTTGGGCTAACCCCGTCAACGCGGTATCGGGGCGAGCGCGTCAACAGACGGCGCGAG 923				

QY 855 CAGCGCTGACCGTGCAGGAGCTCCAGAAATCGTGTGAACGAGCCCTGGC 914
 DB 924 CGCGGCGCTGACCGGCGCCCAACGAGCCGCGAAGAGGGGATCCGACAGGACCTGGC 983
 QY 915 GAGAGCAGGCTGCGCGCGCTCTTCCGTGGTTATGTGAGGACACGCGACGAGGACGAC 974
 DB 984 CAAGCAGCGGCTGACGCGCGACAGCGTGCAGCGCTCGAGGACACGCGACCGGACCCC 1043
 QY 975 GCTTGTGATACCCCATCGAAATCGAATCGTGAATGCGGTTATACGCGCTCGGCGAGATG 1034
 DB 1044 GCTGGGCGAATCGAGGCTCAGGCGCTCTCTCGCACCTACGAGGCGGCGCGGCGGA 1103
 QY 1035 CGCGACCGCGCTGCTGATCGGAGTCGAGTAAGACAACTTGGCACTCTGATATGCTGC 1094
 DB 1104 GGGGAGCGCGCTGAGGCTGGGCTGCTGATGACCTGAGGCGCAACCGATCCGGGCG 1163
 QY 1095 GGGGATCATGAGGCTGCTGAAGTGTCTTGTCTTTCAGACGAGGCGAGATTCTGCGCA 1154
 DB 1164 CGCGCGGCGGCGCTCATCAAGATGTATGCGCATGCGGACGAGGACGCTGCCCGCAC 1223
 QY 1155 CCTCACGCGGAGGCGCTGAACCCCGGATCTCATAGGAGTATCTTCCGCTGACCGTAC 1214
 DB 1224 GCTGACCTCACGAGCGCCACCGCGCGTGAATGCTCCGCGTGAAGTACGAGTGTCT 1283
 QY 1215 GCGCGCCCGGACACCGTGGCGGACTGGAATACGCGCGGAGCGGAGGAGTGTCTGCT 1274
 DB 1284 GACCGAGGCCAGAGACTGCGCGGACACCGGACAGCGCGCGTGGCGGCTGTCTCTT 1343
 QY 1275 CGCGATGAGCGGAGCAACGCGCACGAGTGTGTGAAGAGCGCGCGCGGACGCTGAC 1334
 DB 1344 CGGCGTCAAGCGGACCAAGCGCCATGTATCTTGAAGGCGCGCGCGCGAGAGGCAAC 1403
 QY 1335 ACCGCGC-----GCGCGGAGGAGCGCGGACGAGCTGTGTGTCTG 1376
 DB 1404 GAGCGCGCGCTGCGGAGCTCTCTGCGAGCGCGGAGCGCGCTGCGGTGTGTCTG 1463
 QY 1377 GCGAAGGACCGCGTCAAGCTTGAATGACAGGCGCGCGCTGCGGACCACTTGAAGAC 1436
 DB 1464 CGCGCGGACGAGGCGGCGCTGTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1523
 QY 1437 CTACCTTTCGAGTGTGTGGGAGATGTGGCGCTTCACTGTGGGACGAGCGGACGCGAT 1496
 DB 1524 GCGCGCGGACCTGCGGACCGGCGGAGTGGACACCGCGCTGGGACCAAGCGGCGGCTT 1583
 QY 1497 GAGGACCGGCTCGCGGTGGCGGACGTCGAGGAGGCGGCTGGGACGCGTGGAGCG 1556
 DB 1584 CGACCAAGGCGCGCGCTGTGCGGCGGAGCGTGAAGAACTGTGCGCGGCTTGCAGG 1643
 QY 1557 TGCGGCGGAGGAGACAGAGTGCAGCGCGGCTGGGCGGAGTATCGCGATTCCTCAGCGG 1616
 DB 1644 CTGCGCACCGGAACCGCGCGCGCGGCTGTGTACCGGCGGAGACCGCGGCTCGGCGG 1703
 QY 1617 CAAGCTCGCTTCTTCTTCAACGAGCAGGCGGCGGACGCTGGGATGGGCGTGGGCT 1676
 DB 1704 CAAGGCGCGCTTCTTCTTCAACGAGCAGGCGGACGCGGCGGATGGGCGGAGACT 1763
 QY 1677 GTACGATGATGTGTGCGGCTTCCGAGAGCGTTGACCTGTGCGTGAAGCTTTCAACA 1736
 DB 1764 GAGCGCTACAGCAAGCGGTGTGCGGAGCGCTTGAAGGATGTGCGGCGGCTGACCG 1823
 QY 1737 GAGAGCTGACCGGCGCTCGCGAGGATGATGGGCGGAACGCGGCGAGGCTGACGCGG 1796
 DB 1824 GACACTGACCGGCGCGCTGCGGAGGATGTGTGCGCGCGAGCGGACCGCGGAGCGGCG 1883
 QY 1797 GCTGCTGACGAGACAGCTTTCACCGACCGCGCGCTGTTCACCTTGAATGAGCTGCG 1856
 DB 1884 CTGCTGACAGACAGAGCGCTTACCCAGCGCGCTGTTCGCGGTGAGGCTGCGCTGCT 1943
 QY 1857 CGCGCTGTGCGGCTGTGGGAGTGAAGCGGAGTTGTGCGCGGCGATGAGATGGTGA 1916
 DB 1944 GCGGCTGTGAGAGACTGGGCGTTCGCGCGCGGATGTGTGCGGCGCACTGCTGTGGCGA 2003

QY 1917 GCTGTGTGTGCTGCGGCTGCGGAGGCTGTCTGCTTGAAGACGCGGATTTCTGTGGC 1976
 DB 2004 ACTGACCGCGCTTACCGCGCGCGGAGTGTGTGCTGCTGCGGACCGCTGCGGCTGTGCG 2063
 QY 1977 TGCGCGGCGGCGCTGATGACGAGGCTGTGCGGCGCGGCGGCGGCGGATGTGTGATCGAGG 2036
 DB 2064 CGCGCGGCGGCGGCTGACCAAGGACATGCGCGCGGCGGCGGCGGATGTGTGCGGCGAGG 2123
 QY 2037 GCGCGGCGGATGTGTGCTGTGCGGTGCGGCGGCGGACGAGCGTGTGATGTGCGCGG 2096
 DB 2124 GACCGAGGACGAGGATGCGCGCGGCAACTGCGGAGCGGCGCGCGCGGCGGAGCATGCGCG 2183
 QY 2097 GGTCAACGCTTCGAGACAGGTGTATCGCGGCGCGGCGGCGGACCGCTGACATGCGATGCG 2156
 DB 2184 GGTCAACGAGACCGGAGCGGAGTGTGTGTGCGGCGGAGCGAGGCGCGGCGGCGGCGGCG 2243
 QY 2157 GCGCGGATGAGCGCGCGCGGCGGCGGCGGACCGAAGCGCTCCAGCTTTCGATGCGCTTCCA 2216
 DB 2244 GCGGAGTGTGCGCGCGCGGCGGCGGAGACGAGAGGCTGTGAGGCTCAAGCGCTTCCA 2303
 QY 2217 CTGACCGCTCATGCGCGCGGATGTGAGGCGGCTTGTGCGGCGGATGCGGATGAGCTA 2276
 DB 2304 CTGCGGCGGCTTGAACCGCATGACCGAGGCGGCTTGTGCGGAGTCCACAGAGGAGTGTCTTA 2363
 QY 2277 CGCGCGGCGGCTGATGCTGCTGTGATCAGCAATCTGAGCGGAGAGGCTTTCAGACAGAGG 2336
 DB 2364 CAGCGCGCGGCTTCCGCGGAGTGTCTTCCAGCGCTCACCGGCGGCGCGCGGCGGCGGCGG 2423
 QY 2337 GAGCTCGCGGCGGCTTATGAGTGTGCGGCGGCGGCGGAGGAGTGTGCTTTCGAGATGAGT 2396
 DB 2424 CGGAGGCGGAGACACTGTGTGTGCGGCGGCGGAGAGCGGTGTGCTTTCAGACGAGCGG 2483
 QY 2397 GAAGGCGGTGACGCGGCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2456
 DB 2484 CGCGCGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2543
 QY 2457 GCTGCGGCTGTGCTGCTGCTGATGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2513
 DB 2544 GAGCGCGGCGGCGGCGGAGTGTGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2603
 QY 2514 GCGCGCTGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2573
 DB 2604 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2663
 QY 2574 CGGTGCGGCTGTGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2633
 DB 2664 CGGCGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2723
 QY 2634 CACGTACCTTTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2680
 DB 2724 CACGTACGCTTTCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2770
 RESULT 5
 AAC55840
 ID AAC55840 standard; DNA; 12249 BP.
 XX AAC55840;
 DT 19-JAN-2001 (first entry)
 XX
 DE Complete Mitomycin ORF 1-9 nucleotide sequence.
 XX
 KM Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;
 KM anti-inflammation; immune-enhancer; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW fungicide; pesticide; ds.
 XX
 OS Streptomyces lavendulae.
 XX
 FN WQ200053737-A2.
 XX
 PD 14-SEP-2000.

XX 10-MAR-2000; 2000MO-US06394.
 PF 12-MAR-1999; 99US-0266965.
 PR (MINU) UNIV MINNESOTA.
 PA (SHER) SHERMAN D H.
 PA (MAOY) MAO Y.
 PA (VARO) VAROGLU M.
 PA (HEM) HE M.
 PA (SHEL) SHELTON P C.
 XX
 PI Sherman DH, Mao Y, Varoglu M, He M, Sheldon PC;
 DR WPI, 2000-601980/57.
 XX
 PT Novel nucleic acid molecule comprising mitomycin biosynthetic gene
 PT cluster useful for cloning mitomycin biosynthetic genes for elucidating
 PT the molecular basis of mitosome ring system biosynthesis
 XX
 PS Claim 26; Figure 22; 399pp; English.
 XX
 CC This invention relates to isolated and purified nucleic acid molecules
 CC from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
 CC natural products that contain a variety of functional groups, including
 CC amino benzoguinone and axiridine ring systems. The S. lavendulae
 CC mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
 CC spanning 55kb of DNA. The invention includes an expression cassette
 CC comprising a mitomycin biosynthetic gene operably linked to a promoter,
 CC and host cells transformed with the cassette. The nucleotide, and protein
 CC sequences and the transformed host cells of the invention result in
 CC antitumor, antiinflammatory, cytostatic, immunomodulatory, and
 CC antibiotic activities. The nucleotide sequences are used to elucidate the
 CC molecular basis for the biosynthesis of the mitosome ring system, as well
 CC as to engineer the biosynthesis of novel natural products, e.g.
 CC antibiotics, anti-inflammatory agents, anti-cancer agents,
 CC immune-enhancers, immunosuppressants, agents to treat asthma, chronic
 CC obstructive pulmonary disease as well as other disease involving
 CC respiratory inflammation, or cholesterol-lowering agents or as crop
 CC protection agents (e.g. fungicides or insecticides) as well as
 CC biopolymers, e.g., in packaging or biomedical applications, or to engineer
 CC PHA monomer syntheses. Sequences AAC55782-C55891, AAC55815-C55849 and
 CC AAB32485-832542 represent mitomycin biosynthetic gene cluster DNA
 CC sequences and encoded proteins. Sequences AAC55812-C55814,
 CC AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
 CC cloning of the mitomycin biosynthetic genes.
 CC
 XX Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 other;
 SQ

Query Match 19.1%; Score 814.2; DB 21; Length 12249;
 Best Local Similarity 57.6%; Pred. No. 1e-121;
 Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

18 CGAGCGCGGAGCGGAGATCCGATTCGATCGGAGGAGGAGTCCGCTCCGCGATG 77
 Db 1077 CGAGGACGCGGCGCGGAGCGGATCGGATTCGATCGGAGGAGGAGTCCGCTCCGCGGGA 1136
 Qy 78 CGTGCATCGATCGGAGGAGTTCGAGCGTCTCGAGGAGTCCGCGGACACCGTGGAGG 137
 Db 1137 CGTGCATCGGCGGAGGAGTTCGAGCGTCTCGAGGAGTCCGCGGACACCGTGGAGG 1196
 Qy 138 AGTCCCGCGCGGAGC---GCTGGAGTACACACCGTGTTCGATCCCGACCCCGATGCC 194
 Db 1197 GTTCCCGCGCGGAGC---GCTGGAGTACACACCGTGTTCGATCCCGACCCCGATGCC 1256
 Qy 195 GGGAGAGACGCGCGGATACCGCGCATCTTCTCGAGCGAGTACGCTGTTGACGCGCTC 254
 Db 1257 GGGAGAGAGTACCGCGCGGATACCGCGCATCTTCTCGAGCGAGTACGCTGTTGACGCGCTC 1316
 Qy 255 CTTCTTGGCATCTCGCTCGGAGACGCTGCGGATGACCTTGACATCGACTCTTGTCT 314
 Db 1317 CTTCTTGGCATCGCTCGGAGACGCTGCGGATGACCTTGACATCGACTCTTGTCT 1376

Qy 315 GGAGGTGCTGGAGAGCGCTGAGAGACCGCGCATGCTCCATCGGCGCTGCTGATAC 374
 Db 1377 GGAGGTCTCTGGAGAGCGGCTTTCAGACAGCGCGGCTTCAGCGGATCCTCGGAGGCGGA 1436
 Qy 375 GGAAGCGGAGTGTTCATCGGAGATCGGCGCGCTCGGATATAGAGCGCGCTCGCGAAGC 434
 Db 1437 GGAGGTGCGGCTCTTCGCTGGGCTCCACAGCAAGACTGATCAACGATGCTCGACGC 1496
 Qy 435 GACGCGCTCCGAGAGATCGACGCTCATGCGCGGCTGGGAGAGATGCCACGCTCGAGC 494
 Db 1497 GCGGAGCGTCCGAGGCGCTTC-----ATCGGAGCGGCAACTCCCGCACATCTCTTC 1550
 Qy 495 GGGCGGAATCTGATGCGCTCGGCGTGGAGGCGGCTGTGCGGCGTGAATACGCGCTTA 554
 Db 1551 CGGCGCGCTCGCTACACTTTCGCTTGAAGGCGCGGCGCTGCTCGTGCAGACCGCTTG 1610
 Qy 555 TTCGCTCTGCTGAGGCGCGCTTCACTGAGCTGTCAGAGTTCGCTCGGCGGATGCTC 614
 Db 1611 CTCCTCTCTGCTGAGGCGCGCTGACCTGCGCGGAGTCCCTGGGCGAGGAGGAGTGTCTC 1670
 Qy 615 CACGCGCTGCTGCTGAGGAGTATGCTGATGTTGCTCGCGAGACCTTCGTGTGCTCTC 674
 Db 1671 CCGGCGCTGCGGCGGCGCGCGGAGCGGATGATGCGGACCGCGCTTCATGATGATGAG 1730
 Qy 675 GAAGACCGGCGCGCTGCGCGGAGACGCTGCTGAGGATTTTGGCGGAGCGCGATAG 734
 Db 1731 CCGGACGAGGAGCTGAGCGCGCGCGGAGCGGCTGAGTCTTCTGCGGAGACCGCGGAGC 1790
 Qy 735 GTTGGAGAGGCGAGAGGATGCGCGCTGCTGATGCTCAAGCGGCTCACTGAGACCGCGGC 794
 Db 1791 CACGACCTGATCGGAGGCGCGCGCGGCTGCTGCTGCGCGCTCGGAGCGCGCGC 1850
 Qy 795 GAGCGCGATGAGATTTGCGGCTGATTCGAGAGATCCGCGATCATCATCAAGCGGCGAG 854
 Db 1851 CTTGGGCTAACCCCGGTCAGCGGCTATCCGCGGAGCGCGCTCAACGAGAGCGGCGAG 1910
 Qy 855 CAGCGCTGACCGGCGGAGCGGAGGCTCCCAAGAAATCGTGGAACGCGGCGCTGAC 914
 Db 1911 CGCGGCGCTTACCGCGCGCGCGCGCGGAGCGCGCGGAGCGGCTGATCCGCGAGCGACTGGC 1970
 Qy 915 GAGCGAGGCTGCGCGCGCTTTCGCTGAGGATTAATGTCAGGAGCACAGCGGCGGAGCAG 974
 Db 1971 CAAGCAGCGGCTGAGCGCGCGAGCGGAGCGGCTGAGCGGCTGAGCGGAGCGGAGCGG 2030
 Qy 975 GCTTGTGATCCCGCATCGAAATTCAGAGCTGTGAATGCGGATTAACGCGCTCGGCGAGAT 1034
 Db 2031 GCTGGCGGACCGCATCGAGCGCGCGGCGGCTCTCTGCGCACTTAACGCGGCGGCGGCGGA 2090
 Qy 1035 CGCGACGCGCGCTGCTGATCGGCTGAGAGCAACCTTGAGCATTCCTGATATGCTC 1094
 Db 2091 GGGAGGCGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGG 2150
 Qy 1155 CTTCCAGCGGAGCGGCTGAGAGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1214
 Db 2211 GCTGCACTTCAACGAGGCGGCGGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 2270
 Qy 1215 GCGCGCGCGGAGACCGTGGCGGAGCTGAGATACCGCGGAGCGGCGGAGGAGTACCTGCT 1274
 Db 2271 GACCGAGGCGGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2330
 Qy 1275 CGGATGAGCGGAGCAACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1334
 Db 2331 CGGCGTCAAGCGGAGCAACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2390
 Qy 1335 ACCGCG-----GCGCGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1376
 Db 2391 GAGCGCGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2450
 Qy 1377 GGCAGAGACCGGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1436

[illegible]

Db	3531	GACGCGCGGCA	CGCGATCC	CTGCCCCG	AGCGCCGCCCGA	GACGTTCCG	CCCCGTGCT	3590
Qy	2514	CGCGCGCTGG	CGTACG	ACGCGCGG	AGACCGCTG	CTGAGGCGCT	CGGCGGGCT	2573
Db	3591	GGCGCGCGCG	CGAGCCG	CAACCCG	AGCTCGT	CTACGCGCGT	CGCGAGCGCC	3650
Qy	2574	CGGCGGCGCT	GGATCTC	CGCGGGCG	CGGCGCT	CTTCCCTCA	GGGGGGGGCG	2633
Db	3631	CGGCGCTTCG	CGCGCGCG	GTGGAGAC	CCCCCTG	CTGCCCCA	AGGCCCGG	3710
Qy	2634	CACGTACCTT	GGGAGCG	CGGCGCG	CGCTACTG	AGTACAG	CAAGAACCG	2680
Db	3711	CACGTACG	CGCTTC	CAAGCGCG	CGCCACTACTG	CGCTGGCGGG	CATGCGCG	3757
RESULT 6								
AACS5857								
ID	AACS5857	standard; DNA; 18331	BP.					
AC	AACS5857;							
DT	19-JAN-2001	(first entry)						
DE	Complete nucleotide sequence of the mitomycin gene cluster.							
KM	Mitomycin; biosynthesis; mitosome ring system; antibiotic; anti-cancer;							
KW	anti-inflammatory; immune-enhancer; immunosuppressant; asthma;							
KW	chronic obstructive pulmonary disease; respiratory inflammation;							
KW	fungicide; pesticide; ds.							
OS	Streptomyces lavendulae.							
XX	WO20053737-A2.							
XX	14-SEP-2000.							
XX	10-MAR-2000; 2000MO-US06394.							
XX	12-MAR-1999; 99US-026965.							
PA	(MINU) UNIV MINNESOTA.							
PA	(SHER/) SHERMAN D H.							
PA	(MAOY/) MAO Y.							
PA	(VARO/) VAROGLU M.							
PA	(HEM/) HE M.							
PA	(SHEL/) SHELTON P C.							
PI	Shelman DH, Mao Y, Varoglu M, He M, Sheldon PC;							
DR	WPI; 2000-601980/57.							
PT	Novel nucleic acid molecule comprising mitomycin biosynthetic gene							
PT	cluster useful for cloning mitomycin biosynthetic genes for elucidating							
PT	the molecular basis of mitosome ring system biosynthesis							
XX	Disclosure; Figure 21; 399pp; English.							
XX	This invention relates to isolated and purified nucleic acid molecules							
XX	from the mitomycin biosynthetic gene cluster. Mitomycins are a group of							
XX	natural products that contain a variety of functional groups, including							
XX	amino benzoxiquone and axiridine ring systems. The S. lavendulae							
XX	mitomycin biosynthetic gene cluster comprises 47 mitomycin genes							
XX	spanning 55kb of DNA. The invention includes an expression cassette							
XX	comprising a mitomycin biosynthetic gene operably linked to a promoter,							
XX	and host cells transformed with the cassette. The nucleotide, and protein							
XX	sequences and the transformed host cells of the invention result in							
XX	antitubercular, antiinflammatory, cytostatic, immunomodulatory, and							
XX	antibiotic activities. The nucleotide sequences are used to elucidate the							
XX	molecular basis for the biosynthesis of the mitosome ring system, as well							
XX	as to engineer the biosynthesis of novel natural products, e.g.							
XX	antibiotics, anti-inflammatory agents, anti-cancer agents,							
XX	immune-enhancers, immunosuppressants, agents to treat asthma, chronic							
XX	obstructive pulmonary disease as well as other disease involving							

CC respiratory inflammation, or cholesterol-lowering agents or as crop
CC protection agents (e.g. fungicides or insecticides) as well as
CC biopolymers, e.g., in packaging or biomedical applications, or to engineer
CC PHA monomer synthases. Sequences AAC55782-C55881, AAC5815-C5849 and
CC AAB32485-832542 represent mycoplasma biosynthetic gene cluster DNA
CC sequences and encoded proteins. Sequences AAC55812-C55814,
CC AAC5850-C5886 and AAC5862-C5869 represent PCR primers used in the
CC cloning of the mycoplasma biosynthetic genes.

XX Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 other;

Query Match 19.1%; Score 814.2; DB 21; Length 18331;
Best Local Similarity 57.6%; Pred. No. 1e-121;
Matches 1549; Conservative 0; Mismatches 1108; Indels 30; Gaps 4;

QY 18 CGAGCGCGCAGCCGAGATCCGATTCGATTCGAGCGAGTTCGCTGCGCGG 77
DB 1077 CGAGGACGCGCGCGCGGACCGCGTGCATCGTGGCATGCGTGCCTCCCGGGGA 1136
QY 78 CGTGCATCGATCGAGGCGGCTTCTGACGCTTCTGAGAGGCTCGCGACACCTCGGGCG 137
DB 1137 CGTGCATCGCGGACGACCTGTGCGAGTGTGCGGAGGCGCGGACGCGTCAACGA 1196
QY 138 AGTCCCGCGCGAAC--GCTGGAGATCAGCAGCGTGTGATCCCGACCCGATGCCCC 194
DB 1197 GTTCCCGCGCGACCGGGGCTGGGACGTCAACCGCTCTACGACCCCGAGCCGGACACCC 1256
QY 195 GGGGAGAGCGCCGTTACCGCGCATCTTCTCGAGCGAGTACCTGCTTCAACGCTTC 254
DB 1257 GGGGAGAGCGTACCGCGCGCACCGCGGCTTCTCAAGAGACGCGCGGATTCGACGCGCG 1316
QY 255 CTTCTTCCGATCTCGCTTCGAGAGCGCTGCGGATGACCTCGACATCGACTCTTGT 314
DB 1317 CTTCTTCCGATCTCGCTTCGAGAGCGCTGCGGATGACCTCGACATCGACTCTTGT 1376
QY 315 GAGAGTGTCTGAGAGCGCTGAGAGACCGCGCATCGCTCCATCGCGCGCTGCTCGTAC 374
DB 1377 GAGAGTGTCTGAGAGCGCTTGTGAGAGCGGCGCTTCAACGACACCTTGGGGGCGA 1436
QY 375 GAGAGCGGAGTGTTCATCGGAGTGGCGCGCTCGGATATGAGCGCGCGCTGCGCGAAC 434
DB 1437 GAGAGCGGCGCTTCTGCTGCGCTTCAACGAGACGATCGATCAAGTGTCTGACGCG 1496
QY 435 GACGCGCTCCGAGAGATGACGCTCATGCGGCGCTGCGGAGAGATCCAGGCTCGAGAC 494
DB 1497 GCGGAGCGTCCGAGGCGCTTC-----ATCGGAGCGGCACTCGCGACATCTCTC 1550
QY 495 GGGCGAATCTGTATGCGCTCGGCGTGGAGGCGGCGTGTGCGAGTATACGAGCTA 554
DB 1551 CGGCGCGCTGCTACACTTGGGCTTCAAGGCGCGCGCTGTGCTGTGACACCGCTG 1610
QY 555 TTGCTCTGCTGTGGCGGCTTCACTGTGCTGTGACAGCTTGGCTCGCGGAGATGCTC 614
DB 1611 CTCTCTCTGCTGTGGCGGCTTCACTGTGCTGTGACAGCTTGGCTCGCGGAGATGCTC 1670
QY 615 CACGCGCTGTGTGGTGGTATCGCTGATGTTGTGCGAGACCTCGTGTGTGCTC 674
DB 1671 CTGTGGCTGTGGGCGGCGGCGACGCTGTATGCGACGCTTATGAGTTCG 1730
QY 675 GAGAGCGCGGCGCTGAGGAGCGAGTGTGCAAGGATTTTGGCGGAGCGAGTATG 734
DB 1731 CGGCGAGCGGCGCTTGGCGCGCGCGCGCTGCGAGTCTTCTGCGGACCGCGCGAG 1790
QY 735 GTTTCGAGCGAGCGAGGAGTGTGCTGTGCTTCAAGCGGCTCAAGTGAAGCGCGCG 794
DB 1791 CACCACTGTGTCCGAGGCGCGCGGTGTGCTGTGCGGCTTCTGCGAGCGCGCGCG 1850
QY 795 GAGCGCGATCGATATTTGGCGGTATTTGAGAGATCGCGATCATCAGAGGTGCGAG 854
DB 1851 CTTGGGCTTACCGCGTACGCGGTATCTCGGGGCGAGCGCGCTCAACAGAGCGCGCGAG 1910
QY 855 CAGCGGTGACCGTGCAGACGAGGAGCTCCAGAGAAATGCTGCGAAGCGGCGCTGCG 914

DB 1911 CGCGGCGCTGACCGCGCGCAACGAGACCGCGCGCAACAGCGGATTCGCGAGCACTGGC 1970
QY 915 GAGCGGAGCTGCGCGCGCTTCTGCTGTGTATGTGAGGACACCGCGAGCGGACGAC 974
DB 1971 CACCGCAGCGCTGACCGCGCGACAGGCTGTGAGCGGCTCAGGACACCGCGAGCGACCC 2030
QY 975 GCTTGTATACCCATGCAATTCGAAGCTGTGAATGCGGTATACGCGCTCGGCGAGATGT 1034
DB 2031 GCTGGGCGACCGGATGAGGCGCGCGCGCTTCTGCGACCTACGCGCGCGCGCGCGA 2090
QY 1035 CGCGAGCGCGCTGTGATCGGCTGTGAGAGACCAACTTGGCCATCTGATATGCTC 1094
DB 2091 GGGGAGCGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 2150
QY 1095 GGGGATCACTGAGCTGTGAGAGTGTGTCTTCTTACGACGCGGCGAGATTTCTGCGCA 1154
DB 2151 CGGCGCGCGCGCTGATCAAGATGTATGCGATGCGGACGCGGACGCTGCCCGGAC 2210
QY 1155 CTTCAAGCGGAGCGGCTGAAACCCCGGATCTCATGCGGATGATCTTCCGCTGACCGTAC 1214
DB 2211 GCTGCACTCAAGAGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2270
QY 1215 GCGCGCGCGGACACCGTGGCGGAGTGAATACCGCGCGAGCGCGGAGTGAAGTCTGT 1274
DB 2271 GACCGAGGCGCAGAGCTGGCGGAGACCGGACAGCGCGCGTGTGCGCTGTCTCTT 2330
QY 1275 CGGCACTGAGCGGAGACCAACGCGACGTGTGTGTGAGAGAGCGCGCGGAGCTGAC 1334
DB 2331 CGGCGTCAAGCGGACCAACGCGCTGTGTATCTTGAAGGCGCGCGCGCGGAGAGGAC 2390
QY 1335 ACCGCG-----GCGCGGAGCGACCGGAGAGCTGTGCTGTCTG 1376
DB 2391 GAGCGCGCGCTGCGCGAGCTTCTTCTGACGCGCGGCGCGCGCGCTGCTGCTGTCTTC 2450
QY 1377 GCGCAAGGACCGCGCTGAGCGCTGTGATGACAGCGCGCGCGCGCGCGCACTTGTGAGAC 1436
DB 2451 CGGCGCGGAGGAGGCGCGCTGTGAGGCGGAGCGGCGCGCGCGCGCGCGCGCGCG 2510
QY 1437 CTAACCTTGTGAGTGTCTGAGGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 1496
DB 2511 GCGCGCGACCTGTGCGCGCGCGGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCTT 2570
QY 1497 GAGAGCGGCGCTGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1556
DB 2571 GAGACCGGCG 2630
QY 1557 TCGCGCGGAGGAGACGAGTGTGCGCGGCTGTGCGGAGTATGCGGATTCCTACGCGG 1616
DB 2631 CTTGCGCGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 2690
QY 1617 CAACTGCGCTTCTTCTTCAACCGGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1676
DB 2691 CAAAGCGCGCTTCTTCTTCAACCGGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2750
QY 1677 GTACGATGTATGTGTGCGGCTTCCGAGGCGCTTCACTGTGCGGAGGCTGTTCACCA 1736
DB 2751 GCGCGCTCAAGACAGCGGCTTCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2810
QY 1737 GAGCTTCAGCGCGCGCTTCCGAGGAGTATGAGGCGCGCGCGCGCGCGCGCGCGCG 1796
DB 2811 GCACTTCAGCGCGCGCGCTTCCGAGGAGTATGAGGCGCGCGCGCGCGCGCGCGCG 2870
QY 1797 GCTGCTGACCAAGACGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1856
DB 2871 CTTGCTGACCAAGACGCTTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2930
QY 1857 GCGCGTGTGAGGAGTGTGAGGCGGAGTGTGAGGCGCGCGCGCGCGCGCGCGCGCG 1916
DB 2931 GCGGCTGTGAGGAGCTGTGAGGCGGAGTGTGAGGCGCGCGCGCGCGCGCGCGCG 2990
QY 1917 GCTGTGTGCTGCTGCTGTGAGGCGGAGTGTGCTGTGAGGAGCGGAGTGTCTGTGAGC 1976
DB 2991 ACTGACCGCGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3050

OY	2449	TCAGCGCTGCTCGGCGCTGAGGCGCTGATCGCGAGCG---CCGCGCGCGCGCTGCTC	2505
Db	23453	GGCGTGTCTCAGCGGCTCATGGCCCGGAGAGGCTTCAACGACCCTGCTCCGACAGGACTGCTG	23512
OY	2506	GCATGCTGCGCGCTGAGCGGTGACGAGCCGAGCGACCGTCTCGAGGCGCTTCGCGGCGCTC	2565
Db	23513	CCGACCTCGCGCGCGCGACACCGGCGCCGAGGAACTGACCCTGGTTCACCGCGGTCGCGCGGCGC	23572
OY	2566	TGGCGCGCTCGGTGGCCCTGGTCTCTTGGGCGCGGCTTCTCC---CTCAGGGGGCGGCGG	2622
Db	23573	CACCGCGACCGCGCGCGCGTGTGACTGGAGGGGTGACTTTCGCGACACGACGCGCGCGCCCG	23632
OY	2623	GTGCGCGCTGCGCCAGTACCTCTTGGCAGCGGAGCGGCTACTGGATCGACACGAAAGCCGAC	2682
Db	23633	ACACAGCTGCGCGACCTTAGGGTTTCCAGCGGAGCGGATCTGGCGCCACACACCGCGCGC	23692
OY	2683	GACCGCGCGCTGGCGGACCGCGCTGCTTCGCGAGCGGGTTCAGACGAGGTTCGAGAGGGG	2742
Db	23693	ACGAGCGCGCCACACCGCCCGGATCCGCGCTTGAGCGCGAGTTCTTGGCGCGCGCTCGAGCGG	23752
OY	2743	GGCGCGGTGCGCGCGCGCGGACCGGCGCGAGCGCTCGCTGACCATCGCGCGCCCGAGAC	2802
Db	23753	GACGACGTGCGCGCGCTTCGCGCGCTCCTTGAAGCTTGAGCAGCGCCACCGTACCGCGAGT	23812
OY	2803	GGACCGCGGAGGAGGTTCGAGCGCGCGCGGACCGCTCGTTCCGCGTGGATCGATGAG	2862
Db	23813	GTCCCGCGGCTCACCGCGCTGAGCGCGCGCGCGCGCGAG---GAGACCGAGCTGACCTC	23869
OY	2863	CCAGCGCTGCTTATTCACCTTCGTCTTCGGGTTCACGAGCGCGCGCGCCCTGTGTTCGGC	2922
Db	23870	TGGCGCTTACCGCGCTTACCTTGAAACCGCGCGGCGCGCACCGACCGCGCGCTTCACC	23929
OY	2923	GAGGTTCGAGATTCGCGCGTTCGACGCGCGCGGCGCTCAGCTTCAATATATTCAGCTCGCGCTG	2982
Db	23930	GGCGGCTGCGCTCGGTGCTGTGTGTCCGACAGCACACAGAACCTGTAGAGACAGCGGACCGG	23989
OY	2983	GGCATGTGTCCTCCGACGACTTGCCTGGGAAAGCCCACTTCGCTGCTCTCGGAGCGAG	3042
Db	23990	GCTTGGGCGAGCGGACGTGTGAGACCGCGCTCGGACACACACCTTCGCGGTGACGTGACAC	24049
OY	3043	TGCGCGGCGGCGCATGCTGTGCGCGGCGGAGAGGGGTGAAAGCGGCTCGTGGTGGCCAAACCG	3102
Db	24050	ACCAACCGACGCGCGCGCGCGCTGGCGCGCCGGATTCACCGAACCGCGCG-----GCG	24098
OY	3103	GTCAATCGCCCTTTTCGCGCGGAGCGTTTGCTTACCAACGATCAACACGTCAGCTGCGCTGGTGT	3162
Db	24099	ACCAAGGCGCGGTTTCAAGCGGATGTGCTGTCCGTGTGCGCGTGTCCACCGGAGACCGCGGC	24158
OY	3163	CTGCGCTCGGCTCAGGCGGCTCTTGCGGCGATGAGGCGGCGCGCATGCGCCGTCCGCTTACTTG	3222
Db	24159	ACCCCGGTGGCGCCCGCGCGCTTCAACCTTCACACACACCGCGGTCCAGGCGCTTCGCGGACG	24218
OY	3223	ACGCGCATGTATACGCGGCTCGACAGAAATGCGCGCTTCAGCGCGGAGAGCGGGGTGATTC	3282
Db	24219	CCGCGCATTCGACGCGCGCTGTGTGAAAGTCAACCGCGGAGCCGTGTGCGCGCGCGCG	24278
OY	3283	CATCGCGCGACCGCGGCGGTTCTGTTCGCGCGGTGCAGTGGCGGACGACGATGGAGGCC	3342
Db	24279	AACAGGTTCACCGCGCCCGGAAACAGGCGCGCGCTCTTGGGCGCTTGGCGCGCGCTTCGCTTG	24338
OY	3343	GAGGTTCATGCG--GAGGCGCGGACACCGCGCGGAAAGCGCGCTTACTGTGAGTGTCTGGGGG	3400
Db	24339	AATCTGCGCGCGCGGTTTGGCGGACCTTCGACTGCGCCGACCTTGTGAACCGCGCAAGCCG	24398
OY	3401	TGCGGTATGTGAGCGAATTCCGCTTCGGAACCGGATTCTGTGCGCGGACGTCGCGCGGTGGACG	3460
Db	24399	CCCGCGGTTTGGCGCGGATGCTGCGCGCTTACGACGCGCGAGACGCGGATGTGCGCGCGC	24458
OY	3461	GGCGCGAGGAGTATAGACTGTGTCTCAATCTCGCTTTCGGCGAGCTGATTCACAGATTT	3520
Db	24459	CTTCCGCGGTTCTCTTCGCGCGCTTGGCCACGCGCCCGCGCGCGCCCGACACCGCGCGCA	24518

ID	Accession	Gene Name	Location/Qualifiers	Length (bp)
OY	3521	TCATCTCCGCGCATCGCAAGCGCGGTTTGAGAGCTCGGCAAGCGCACTGTAAAGCGG		3580
Db	24519	CCGCTTCAGACCCGCGGCGCGGACCGCTCTATATACCGCGGACCGCGCGCATCGCGG		24578
OY	3581	ATTAACGACCTCGGAGCTCGGCGCGCTTCCTGCGCAATCTCTTTCGCTGTGATCTCC		3640
Db	24579	GCCACGTGTGCGCCGCGCGCTGTGGCCCGCGGACCGGCGGACCACTGCTGTACACGCGCGC		24638
OY	3641	GGGGGATGATGCTCTGACGCGCGCGCGCGGATCCGTGCGCTTTTGAGAGAGCTCTTGCGCC		3700
Db	24639	GCGGCCCCGCGCGCGCGCGCGGAGCGGCTCCGCGCGCACTGGAGAACTGGGCGGCC		24698
OY	3701	TGATCGCGGAGGAGGAGTTCACCCCTCCCGCCATCGGAGCGCT		3743
Db	24699	GGGTACCTCTGCGCGCTGCGATCGGCGGACCGGACCGGAGCGGCT		24741
RESULT 8				
AD	AD17186	standard; DNA; 125401 BP.		
XX	AD17186;			
XX	29-NOV-2001	(first entry)		
XX	Streptomyces noursei	nystatin PKS gene cluster DNA.		
XX	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster; antifungal; antibiotic; de.			
XX	Streptomyces noursei.			
XX	Key	Location/Qualifiers		
FT	CDS	6337..34771		
FT		/*tag= a		
FT		/product= "NysI complete protein"		
FT	CDS	34792..51099		
FT		/*tag= b		
FT		/product= "NysJ protein"		
FT	CDS	51155..57355		
FT		/*tag= c		
FT		/product= "NysK protein"		
FT	CDS	57503..58687		
FT		/*tag= d		
FT		/product= "NysL protein"		
FT	CDS	complement (58786..58980)		
FT		/*tag= e		
FT		/product= "NysM protein"		
FT	CDS	/note= "CDS does not include start codon"		
FT		complement (59045..60241)		
FT		/*tag= f		
FT		/product= "NysN protein"		
FT	CDS	/note= "CDS does not include start codon"		
FT		complement (60238..61296)		
FT		/*tag= g		
FT		/product= "NysO2 complete protein"		
FT	CDS	120628..121308		
FT		/*tag= h		
FT		/product= "NysR4 (long) protein"		
XX	MO200159126-A2.			
XX	16-AUG-2001.			
XX	08-FEB-2001; 2001WO-GB00509.			
XX	08-FEB-2000; 2000GB-0002840.			
XX	10-APR-2000; 2000GB-0008786.			
XX	14-APR-2000; 2000GB-0009387.			
XX	(UNO-) UNIV NORGES TEKONISK NATURVITSENSKAPSELIGE.			
XX	(SNTF) SINTEF STIETELSEN IND TEK FORSK.			
XX	(ALPH-) ALPHARMA AS.			

PA (SINV-) SIVENT AS.
PA (DZIE/) DZIEBEMSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVILK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI: 2001-557614/62.
DR P-PEDB: AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR AAE10149, AAE10150.
XX
XX New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
PS Claim 1, Page 188-254, 266pp; English.
XX
XX The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrocyclic antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

Query Match 18.9%; Score 805; DB 22; Length 125401;
Best Local Similarity 53.0%; Pred. No. 2,7e-120;
Matches 1996; Conservative 0; Mismatches 1715; Indels 52; Gaps 11;

QY 4 GCGGATCGTCCATGAGCGCGCAGCCGAGATCCGATTCGATCGTGGACGAGTTCC 63
DB 21008 GCCGTGCCACCGCGCGCCCGACGCGAGACCCCGATCGATCGTCCGATGAGCTGC 21067
QY 64 CGTCTCGCGGCGGCGTGCATGATCGAGCGGATTCGAGCGCTCTCGAGGCGTGGCG 123
DB 21068 CCTTACCCCGGCGGCGGCGTCCCGCGAGACCTGTGGCGATCCCTTGAACGAGTTC 21127
QY 124 GACACCGTGGGCGAGTCCCGCGGAC---GCTGGGATGAGCAGCGTGTGATCC 180
DB 21128 GACGCGATCTCGGCGTCCCGCGACGCGGCGTGGAGCGCGAGGCGCTTACGACCGG 21187
QY 181 GACCCCGATCCCGCGGAGAGCGCCGTTACGCGCGCATCTTCTCTGAGCGATGATCC 240
DB 21188 GACCCCGACCGGCGCGCGCATCTCTCGTCCAGGCGGATCTCTGCGCAGCGTCCG 21247
QY 241 TGCCTCGACCGCTCTTCTGCGCATCTGCGTCCGCGAGCGCTGCGGATGAGCCCTGCA 300
DB 21248 GAGTTGACCGCGGCGTCTTCTGCGCATCTGCGCGGAGCGCTGTGAGTGAACCGCAG 21307
QY 301 CATCGACTTGTGTGAGGAGTGTGCTGGAGCGCTGGAGAGCGCGCGATTCGCTCCATCG 360
DB 21308 CAGCGGCTCTCTGTGAGAGCCCGCTGGAGGCGTTGACACGCGCGGATTCGACCGCGTC 21367
QY 361 GCGCTCTGTGTGAGAGAGAGAGTGTTCATGCGGATGCGCCCGTCCGATATGAGGCC 420
DB 21368 GCGCGAGCGCGAGCGCGACCGCGACCTTGTGTGCGCGCATACGAGGATCAACGCTCC 21427
QY 421 GCGCTGCGCGACGAGCGCGGTCGAGAGATGACGCTCATGCGCGGCGTGGGAGAGT 480
DB 21428 GCGCTGCGCGACGAGCGGCTC-----GAGAGCGCATGATCACCGCGAGCTTC 21478
QY 481 CCCAGCGTGGAGCGCGCGAATCTGATGCTTGGCTTGGGCTGGAGGCGGTGTGTGCG 540
DB 21479 TCCAGTGTCTGTCCGCGCGGAGTCTTACCTTGTGCTTGAAGGCGCCCGCGGTCAAG 21538
QY 541 GTGAGTACGCGCTATTCTGTCTGTGCTGTGCGCGCTTCACTGCGCTGTGAGAGCTTGGCG 600
DB 21539 CTTCACACCGCGCTCTCTCTCTCCGTGTGCGCATCACTGGCGCTTCCGACCGCTCCGCG 21598
QY 601 TCCGCGGAGATGCTCAACGCGCGCTGAGTGGGATTCGATATGTTGTGCGCAGACACC 660

DB 21599 AACGGGAGACTCGCTGCGCTTGCGCGCGCGCGCTGACGATCATGTCCACCCCGATTCG 21658
QY 661 CTCTGTGCTCTTCGAGAACCCCGGCGCTGCGCAGAGAGCGGTCTGTCAAGCATTTTCG 720
DB 21659 TTCTGTGCTCTCAGCGCGCAGCGCGCTCCCGAGGACGCGCGCTGTCAAGCGTACCGG 21718
QY 721 GCGAGGCGGATGAGTTCGAGCAGAGGCGAAGGATGCGCGGTGTGTCTCAAGCGCTC 780
DB 21719 GACGCGCGCGAGGATACCTTCGCGAGGCGCTGCGCTGTGTCTGTGAGCGGCTG 21778
QY 781 AGTGAAGCCCCGCGCGAGCGGATTCGATATTTGCGGATTCGAGATCCCGCATCAAT 840
DB 21779 TCCAGCGCGCGCGCAGAGGCGACAGGATGCTCGCGGTGATCCGCGCTCCGCGTCAAC 21838
QY 841 CACGAGGTGCGAGCAGCGGTCTGACCGTGTCCGACGAGGACCTCCCAAGAAATGTCGTG 900
DB 21839 CAGAGCGCGCTCCCAACGCGCTGACCGCACCCAGCGCGCGTCCACAGCGCGTCAATC 21898
QY 901 AACCGGCGCTTGGGAGAGCGAGCTGCGCGCGCGCTTCTCGGTGATATGAGGCAAC 960
DB 21899 CGCGAGGCGTGGCGCACTCGCGGTGCGCGCGCGCATGACGCTCTGAGAGGCGCAC 21958
QY 961 GCGACGCGCGACGCGCTTGTGACCCCATCGAAATCCAGCTTGAATGCGGTATACGCGC 1020
DB 21959 GCGACCGGTACCGCGCTTGGGAGACCCCATCGAGGCGAGCGCGCTCTGCGCACCTACGCG 22018
QY 1021 CTGCGGCGAGATGTCGCCACCGCGCTGTGATCGGCTGTGAGAACCAACCTTGGCAT 1080
DB 22019 CAGAGCGCGCGCGCGCAACGCGCGCGCTGCTGCGCTGTGAGTCCCAACATCGCGCAC 22078
QY 1081 CCGATGATGCGGCGGAGTCACTGAGGCTGCGTGAAGGTGCTGTCTTCAAGACGGG 1140
DB 22079 ACCGAGTGGATCTCGGCGTGTGACGCTCATCAAGCTGTGTCCGCGCTTCAAGAGGC 22138
QY 1141 CAGATTCTGTGCACTTCACGCGCGAGCGCTGAACCCCGGATCTCATGAGGATGATCTT 1200
DB 22139 GTGTGTCCCAAGTCCCTGCAATGACGCGCGCTTCAACCGATGATGATGATGATGATG 22198
QY 1201 CGGCTGACGTCACGCGCGCGCGCGGACACCTGTGGCGCATGGAATACCGCGAGCGGCG 1260
DB 22199 GCGATCGGCGTGTCTACCGAGACGACCCCGTGGCGCGAGACCGCGCGCGCGCGCGCC 22258
QY 1261 GGGGTGACCTGTTCTGAGATGAGCGGAGCAACCGCGCGAGTGTGTGAAGAGGCGCG 1320
DB 22259 GCGGTCTCTCTTGTGAGATGAGGAGCAACGTCACACCATCTTGAACAGGCGCCC 22318
QY 1321 GCGGCGAGTGCACACCGCGCGCGCGCGAGGACCGCGAGAGCTGCTG-----GTGCTG 1374
DB 22319 GCGAGCAGGCG 22378
QY 1375 TCGGCAAGAACCGCGCTGACCCCTGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCG 1434
DB 22379 TCCGCGCGCGCGAGGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 22438
QY 1435 ACCTACCTTGTGCACTGTGTGGGAGATGAGGCTTCACTGTGAGAGCGACGCGAGCGG 1494
DB 22439 GAGCGCGCGAGGCG 22498
QY 1495 ATGAGCAGACCGCGCTTGTGAGCGCGCGCGAGTGTGAGGAGGAGGCTGCGCGAGCCCTGAGC 1554
DB 22499 CTGGAACCG 22558
QY 1555 GCTGCGCGCGAGAGCAGACGTCGCGCGGTGTGCGCGAGTATGCGCGATTCCTCAAGC 1614
DB 22559 GCGCTGTCTGACG 22612
QY 1615 GCGAAGCTCGCTTCTTTCACCGGAGAGGAGGCGAGAGCGTGGGAGTGGGCGCGTGGG 1674
DB 22613 GCGCGGACCGCTTCTGTTCACCGAGAGGAGGCGAGCGCGCGCGCGCGCGCGCGCGGAA 22672
QY 1675 CTGTACGATGTATGCTCGCGCTTCCGCGAGGCGTTCGACCTGTGTGCGTGAAGCTGTTCAC 1734

Dh 22673 CTCCACGCGCTACCCGGTGTGCGGACGCGCTGACGAGGTGTGGCCCGGCTCGAC 22732
Qy 1735 CAGAGCTCGACCCGCGCCTCCGCGAGGTGATGTGGCGGAACCGGCGAGGCTGACGCC 1794
Db 22733 GACGGACCGGACCCGGCGCTGCGGAGTCTGTTCGCGCGCGGACTCGGCGAGGCC 22792
Qy 1795 GCGGTGCTGACACGACGACCTTCAACCGAGCGCGGCTGTTCACCTTGAATATGCGCTC 1854
Db 22793 GCGCTCTGACACGAGACCGGCTACGCGCAGCGCGGCTGTTCGCGGTGAGTGTGCGCTG 22852
Qy 1855 GCCCGCGTGTGGCGGTGTGGGGGTGTAGAGCGGAGTGTGTGCGCGGCTATAGCATGTGCT 1914
Db 22853 TTCCGCTGTGACGCTCTGAGGCGCTGACCCCGGACTACCTGTGCGCGGCACTCGTCCG 22912
Qy 1915 GAGCTGTGTGCTGCTGCTGTGGCGGCGCTGTTCCTGTGAGAGACGCGGTTCCTGTG 1974
Db 22913 GAATCTGCGCGCGCGGACGCTGCGCGGCTGTGTGCTGTGAGACGACGCTGCACTGTGTC 22972
Qy 1975 GCTGCGCGGCGCGCTGTGATGACGCGCTGCGCGCGGCGGAGCGAGTGTGATCGAG 2034
Db 22973 GCCCGCGCGCGGCTCATGACGCGCTGCGCGGCGGCGGAGTGTGCGCTCGAG 23032
Qy 2035 GCGCGGAGGCGGAGTGTGAGTGTGCTGCGGCTGCGCGCGGCGGCGGCGGCTGCGTCC 2094
Db 23033 GCCCGGAGGACGAGGTCTGCGCTCTGAGAGGCGCTCACGACCGGAGTGTCTGCTGCC 23092
Qy 2095 GCGGTCAACGCTCCGCGACAGAGTGTGATCGCGGCGCGGCGGCGGCGGCGGCGGCTG 2154
Db 23093 GCCGTCAACGCGCGGCGGCTGCTGTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCT 23152
Qy 2155 GCGCGCGGAGTGTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2214
Db 23153 GCCGACCTTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 23212
Qy 2215 CACTTCACGCTCATGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2274
Db 23213 CACTTCGCGCTGATGAGCGCATGTCTGACGACTTCGCGCGGCGGCGGCGGCGGCTGAC 23272
Qy 2275 TACCGGCGGCGGCTGATGTCTGTGTCAGCATCTGA---GCGGAGAGGCTTGGACAGAC 2331
Db 23273 TACCAACCGCGGAGATTCGCTGTGTGAACTGACGCGGCGGCGGCGGCGGCGGCGG 23332
Qy 2332 GAGGTGAGCTGCGCGGCTATGTGGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCTG 2391
Db 23333 CAGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 23392
Qy 2392 GGAATGAGAGCGCTGACGC---GCGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCG 2448
Db 23393 GGCATTCGACTGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 23452
Qy 2449 TCGACGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2505
Db 23453 GCGGTGCTCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 23512
Qy 2506 GCATGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2565
Db 23513 CCGACCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23572
Qy 2566 TGGGCGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2622
Db 23573 CACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23632
Qy 2623 GTGCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2682
Db 23633 ACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23692
Qy 2683 GACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2742
Db 23693 ACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23752
Qy 2743 GCGCGGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2802
Db 23753 GACGACGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23812

Qy 2803 GAGCGCGGAGAGGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2862
Db 23813 GTCCCGCGGCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23869
Qy 2863 CAGGCGGTGCTGATCACTGTGTGCTTGTGAGTCAAGAGCGGCGGCGGCGGCGGCGG 2922
Db 23870 TGGCGTAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 23929
Qy 2923 GAGGTGAGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2982
Db 23930 GCGCGTGTGCTGT 23989
Qy 2983 GGCATGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3042
Db 23990 GCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24049
Qy 3043 TGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3102
Db 24050 ACCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24098
Qy 3103 GTATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3162
Db 24099 ACCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24158
Qy 3163 CTGCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3222
Db 24159 ACCCGGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24218
Qy 3223 ACCGATGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3282
Db 24219 CCGGATGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24278
Qy 3283 CATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3342
Db 24279 AACAGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24338
Qy 3343 GAGGTTCATCG---GACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3400
Db 24339 AACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24398
Qy 3401 TGCCTATGTGAGGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3460
Db 24399 CCGCGCGGTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24458
Qy 3461 GCGCGGAGGAGTGTGAGTGTGTGCACTGTGTGTGTGTGTGTGTGTGTGTGTGT 3520
Db 24459 CTTCCGCGGCTTCTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24518
Qy 3521 TCAATCTCTGCGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3580
Db 24519 CCGGCTTGTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24578
Qy 3581 ATTAACAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3640
Db 24579 GCGACGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24638
Qy 3641 GGGGAGATGTGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3700
Db 24639 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24698
Qy 3701 TGAATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3743
Db 24699 GGGTACCTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 24741

RESULT 9
AA517367
ID AA517367 standard; DNA; 33529 BP.
XX AA517367;
AC
XX
DT 12-MAR-2002 (first entry)

OY	1343	CGCCGAGACGACCGGACGACCTGTGGATGTGTCCGGAAGGAAACCGGCTAGCCCTGGAAG	1402
Db	26827	AGCGGCGGCGCTGCCTATTTCCGCTGGTCTGTCCGGTGGACAGAGCCGCGGTGAATG	26886
OY	1403	CACAGCGGCGCGGCTGCGGACCAATCTGGAGACTTAACTTCGAGTGTCTTGGGCGATG	1462
Db	26587	CGCAGGCGGGGCGGTGGGCGAAGTGGCTGGAAAGACAACGGAGGAGTGGGATGGTCCGACG	26746
OY	1463	TGGGCTTCAGTCTGTGCGACACACCGCGACGCGCATTTGAGACACCGGCTCGGAGTGGCGGCA	1522
Db	26747	TGGTGCAGACGGCGGCGCTGCACCGACGCACTTCAGATTCGCGGCGGTGGCTTTCGCGG	26806
OY	1523	CGTGGAGGAGAGGAGCTTCGGGCGACGCTTCGACGCTGGGCGGACGAGGACAGAGCTGCGCGG	1582
Db	26807	CGAGCGGCTGGGAGGAGCTGTGGAGGGTCTTTCGCGCGCTGTCTGTGGGGCGCGGATATCGCG	26866
OY	1583	GTGCGGTGGCGCAGATATCGCGAATCTTCACGCGGCAAGCTCGCTTTCCTTACCGGAC	1642
Db	26867	CGGTGTGTGAGGCGGAGACCGGCATA--GCGAGGCGGGAAAGCTTGGCGTGTCTGTTCACGGGCG	26923
OY	1643	AGGAGGCGCAGACGCTGAGGATGGGCGGTGGGCGTGAACATGTATGATGTCGCGGTTCGCGG	1702
Db	26524	AGGCGACGACGCGCTCGGAGTGGGGAAGGCTTTTACAGATGTACCCGTTGTCCGTG	26983
OY	1703	AGGCGTTTCAACTGTGTGCTGAGGCTGTTCACACGAGAGCTTCACCGGCGCTTCGCGAGG	1762
Db	26984	CGGCGTTTCAGACAGATGTGGGAGGCGCTGGACCGGACTTCGACCGTGGGTTGAGAGAGG	27043
OY	1763	TGAATGTGGGCGGAAACCGGCGACGCTGCAGCGCGCTGTCCACAGACAGCTTTCACCC	1822
Db	27044	TGGTGTTCGCGGCGCGCGGAGCGAGGAAGAAACGACGCTGGAGCGGAGCGAGTTCACGCG	27103
OY	1823	AGCGGCGGCTGTTCACCTTCGAATATGCGCTCGCGCGCTGTGGCGGCTGTGGGAGTATG	1882
Db	27104	AGCCCGGAGCTGTTCGCTGCTGGAAGTGGCGGTGAACGTCAGTGGAGTGTGGGGGCTGA	27163
OY	1883	AGCGGAGATTGTGTGCGGCGGCATATGACATCGGTGAGCTGTGGTGTGCTGTGCGGCGG	1942
Db	27164	AGCCCGCTGCGCTTCTTGGGCGACTCGATTAAGGAAGCTGAACGCTGTGCGACGTGGCGGGG	27223
OY	1943	TGTTCTGCTTGAAGACGCGGATGTTCCTGTGTGCTGTGCGCGCGGCGGCTGATGACAGCGCG	2002
Db	27224	TGCTGAGACTTTGCGGAGACGACGAGAGCTATGTGTGCGCGCGGTCCGCTGAATGCAAGGAT	27283
OY	2003	TGCCGCGCGGCGGAGCGATGTGTGCATTCGAGGCGCGCGGAGGCGCATGTGGCTGTGCGG	2062
Db	27284	GCGAGGCGCGGAGGAGCGATGTGTGCTGTGAGGCGCTTCGAGGCGGAGGTGCAACGCGGCGCG	27343
OY	2063	TGGGCGCGGACGACAGCTCGG-----TGTGCATTCGCGCGGTTCACGCTCCGAGACGAG	2116
Db	27344	TGTGTGAGAGTGTGGGGGCGACAGGCGGACTGTGACATTCGCGGGGCTGAAACGCGCGATGACA	27403
OY	2117	TGCTCATTCGCGGCGCGCGGCAACCCGTGATCGATTCGCGGCGGCGGATATGCGCGCGCGG	2176
Db	27404	CGGTGTGAGCGGCGGGAAGAACCGCGGTCTCTCCGCTGTGGCGGACAGCGCTGAGAGCGCAGG	27463
OY	2177	GGGCGCGGACCAAGAGCGCTTCAGCTTTCGATGTGCTTCCACTACCGCTCATGCGCCGGA	2236
Db	27464	GCCGCGCCACCGCGCGCTCTCGTGTGTGTGCGACCGCTTTCACACCGCGCACATGACCGGGA	27523
OY	2237	TGCTGAGAGGCGTTTGGGCGTGTGGCGGAGTCGAGTGAAGCTACCGGCGGCGGTGTGATCTTC	2296
Db	27524	TGCTGTGAGAGTTCGGGAAGGTGTGCGGCGGGAATGTGCATATACGCGCGGCGGACTGTGCGG	27583
OY	2297	TGCTCAGCAATCTGA--GCGGGAAGCTTTCACAGACGAGGTGAGCTCGCGGGCTATT	2353
Db	27584	TGCTGTGAGCGGCGTGTGACCGGCGGAGCTCGGTGTGCGAAGAAAGCGCTGATGTGCGCGGAGTACT	27643
OY	2354	GGGTGGCCCAACGCGCGAGAGGTGTGTGCGCTTCGCGATGAGAGGAAGGCGTGTGACCGCGG	2413
Db	27644	GGGTGAGGCAAGTGTGCGAGAGCGGTGTGCTTCTGTGACGGGAATGCGCAAGCTTTCGCGCGG	27703

Oy	2414	CCGGTGGGGGACCTTCGTGAAAGTGGATCCGAAATGACAGCTGCTGGCCCTGGTGCCTGG	2473
Db	27704	CGGGGGTGAACACATATAGTTCGATGTGGGCGGATGGCGTCTGTGACGCGCTGGGGGCGG	27765
Oy	2474	CTTCGATGCCGGAACGCCCGGCGCGCTGCTGCATC-----GTCCGCGCTGGGGCTGG	2527
Db	27764	GGTCCCTGGCCGGAGGGAGCGGAGGACGCTTTGTGGCGAGCCTGGCGGAGAGCGAAGG	27822
Oy	2528	ACGAGCGGAGCAGACCGTGTCTGAGAGCGCTCGGCGGCTCTGGGGCCGTGGGTGGCTGTCT	2587
Db	27824	AAGAGCGCGGCTGGAGCGAGCGCGGTGGCGACATGTACATGTCAGGGGACGAGAGTGAAT	27883
Oy	2588	CTTGGGCGGGCTCTTCCCTCAGGGGGGGGGGGGTGCCCTGGCCCATGTAACCTTTGGC	2647
Db	27884	GGGCGCCAGGTGTGTGCGGGCCGTGGCGGCGCGGCCCTGTGAAGCTGCGGAGTGTACGCGTTCC	27943
Oy	2648	AGCGCGAGCGCTACTGGATGCAACAGAAAGCGACGACCGCGGCGCTGGCGAACCGCGCTGG	2707
Db	27944	AGCGCGAGCGCTACTGGCTGTGAAAGCCCGGAAGCGGTACCGACGTGGGCTTCGGCGGGCT	28003
Oy	2708	CTCCGGAGAGGGATCAGACGAGAGTGCAGAGAGGGGGGCGCGGTGCGCGCGGACCGAGC	2767
Db	28004	TGAGGAGAGTGGGGGCAATCCGCTGCTCGAGACGGCCAAAGAACTGGCGCGACCGGACGCGC	28063

CC	RESULT 10
AAAF24892	ID AAF24892 standard; DNA; 20394 BP.
XX	
AC	AAF24892;
XX	
DT	20-APR-2001 (first entry)
DE	Pimaricin biosynthesis associated polyketide synthase gene.
XX	
KW	polyketide synthase; oxidative modification; metabolite; antibiotic;
RN	anticancer; pimaricin; ss.
XX	
OS	Streptomyces natalensis.
XX	
FH	Key Location/Qualifiers
FT	1..20394
CDS	/*tag= a /product= "polyketide synthase"
XX	
PN	MO200077222-A1.
XX	
PD	21-DEC-2000.
XX	
PF	14-JUN-2000; 2000MO-BP6227.
XX	
PR	14-JUN-1999; 99EP-0201893.
XX	
PA	(STAM) DSM NV.
PI	Martin JF, Aparicio JF, Collina AJ;
DR	WPI.; 2001-080693/09.
DR	P-PDB; AAB31558.
XX	
PT	New polynucleotides encoding enzymes involved in the biosynthesis of
PT	pimaricin, useful for modifying the biosynthesis of pimaricin and in
XX	the synthesis of new compounds -
PS	Disclosure; Page 53-80; 116pp; English.
XX	
CC	The present sequence encodes a polyketide synthase which is associated
CC	with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC	is useful for the oxidative modification of a methyl group of a suitable
CC	compound, e.g. a bioactive compound including a secondary metabolite,
CC	antibiotics and anticancer agents. Recombinant cells comprising the
CC	gene are useful for the production of pimaricin. The polyketide synthase
CC	polynucleotide may be over expressed in Streptomyces, leading to an

CC increase in the biosynthesis of pimarinin, as a source of primers for
CC amplification reaction and as probes.

XX Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

Query Match 18.0%; Score 766.6; DB 22; Length 20394;
Best Local Similarity 53.6%; Pred. No. 4,1e-114;
Matches 1644; Conservative 0; Mismatches 1549; Indels 46; Gaps 10;

```
QY 33 AGATCCGATTCGATTCGAGCCGATTCGCGTCTGCGGCGGATGATCGATCGAG 92
DB AGAGCCCTCGGATTCGATTCGATTCGCGTCTGCGGCGGATGATCGATCGAG 158
QY 93 CGGCTTCGAGAGCTTCGAGAGGCTCGGCGGATTCGCGTCTGCGGCGGATTCG 151
DB 159 GAGCTGAGAGCTTCGAGAGGCTCGGCGGATTCGCGTCTGCGGCGGATTCG 218
QY 152 -GCTGAGAGCTTCGAGAGGCTCGGCGGATTCGCGTCTGCGGCGGATTCG 209
DB 219 CGGCTTCGAGAGCTTCGAGAGGCTCGGCGGATTCGCGTCTGCGGCGGATTCG 278
QY 210 TACGCGGATTCGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 269
DB 279 GCGGAGGCGGATTCGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 338
QY 270 GCGTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 329
DB 339 CGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 398
QY 330 GCGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 389
DB 399 AGCCGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 458
QY 390 CATCGGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 449
DB 459 TGTGCGGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAG 512
QY 450 GATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 509
DB 513 GGTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 572
QY 510 TGCCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 569
DB 573 TGTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 632
QY 570 GCGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 629
DB 633 CGCCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 692
QY 630 TGGGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 689
DB 693 CGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 752
QY 690 GCGGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 749
DB 753 TGCCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 812
QY 750 AGGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 809
DB 813 GGTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 872
QY 810 ATTGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 869
DB 873 CCGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 932
QY 870 GCGGATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 929
DB 933 CGCCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 992
QY 930 GCGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 989
DB 993 CGCCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1052
```

```
QY 990 GGAATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1049
DB 1053 CGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1112
QY 1050 GATTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1109
DB 1113 ACTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1172
QY 1110 GCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1169
DB 1173 GATGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1232
QY 1170 GCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1229
DB 1233 GCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1292
QY 1230 GCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1289
DB 1293 CTGCTGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1352
QY 1290 CAAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 1340
DB 1353 CAAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 1412
QY 1341 GCGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 1400
DB 1413 GCGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 1472
QY 1401 TGCAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 1460
DB 1473 TGCAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAG 1532
QY 1461 TGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1520
DB 1533 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1592
QY 1521 GAGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1580
DB 1593 GAGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1652
QY 1581 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1640
DB 1653 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1706
QY 1641 ACAAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1700
DB 1707 GAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1766
QY 1701 GAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1760
DB 1767 GAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1826
QY 1761 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGG 1820
DB 1827 CATGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1886
QY 1821 CAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1880
DB 1887 CAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1946
QY 1881 AGAGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 1940
DB 1947 GAGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 2006
QY 1941 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 2000
DB 2007 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 2066
QY 2001 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 2060
DB 2067 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 2126
QY 2061 GGTGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCTTCGAGAGGCT 2120
```


PT DNA encoding Streptomyces fradiae tylactone synthase domain - for
production of tylosin-related polyketide compounds

PS Claim 2; Pages 8-66; 220p; English.

XX This sequence represents the tylactone synthase gene cluster of the
CC invention. This sequence is also referred to as the tylG gene, and was
CC isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC proteins which direct the synthesis of the polyketide tylactone, isolated
CC from Streptomyces fradiae. Tylactone is the basic building block of the
CC antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC type of carboxylic acids incorporated, the number of carboxylic acids
CC incorporated and/or the post-condensation reactions performed, thereby
CC resulting in novel tylosin-related polyketides.

SQ Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

Query Match 17.7%; Score 755.2; DB 18; Length 43280;
Beet Local Similarity 56.0%; Pred. No. 2,6e-112;
Matches 1603; Conservative 0; Mismatches 1208; Indels 53; Gaps 7;

QY 24 CGCAGCCGGAAGATCCGATTCGATCGTCGAGCCGAGTTCCTGCGCCGATGCGGTAT 83
Db 36350 CGCGGCGCGGAGCGGATCGCGTGTGCGATGCGCTCCGCTACCGGGGCGATGTCAC 36409
QY 84 CGATCGAGCGGGTTCTGACGCTCTCGAGGGCTTCGCGACACCGTGGGCGAGTCCC 143
Db 36410 CGGCCCCGAGAGCTGTGGGACCTGGTGGCCGAGGCGGAGCGGATGGGCGCTTCCC 36469
QY 144 CGCCGAC--GCTGGAGTACGACGCGTGTGATCCGACCCCGATGCCCCGAGGGA 200
Db 36470 CTTGACCGGGGCTGGAGCTGGCGTGGTGAACGACCGGATCCGATCCGAGGAC 36529
QY 201 GACGCCCCGTTACGCGCGCATCTTCTCGAGCGACGTAGCTCTTGACGCTCTTCTT 260
Db 36530 CACGTACTCCGGAGGCGGGGTTCTGGAGGCGCGCGGTGACCTTCGACGCGCTTCTT 36589
QY 261 CGGCACTTCCTCGCGAAGCGCTCGGATGACCTTCGACATCGACTTTCTGAGAGT 320
Db 36590 CGGCACTTCCTCGCGAAGCGCTCGGATGACCTTCGACATCGACTTTCTGAGAGT 36649
QY 321 GTGCTGGAGGCGCTGAGAGACGCGCGATCGCTCCATCGGCGCTGTCGTATCGGAAAC 380
Db 36650 GTCTTGGAGGCGCTGAGAGACGCGCGATCGCTCCATCGGCGCTGTCGTATCGGAAAC 36709
QY 381 GGGAGTTCATCGGAGTCCGCGCTCGGATATGAGGCGCGCTGCGCGAAGCGAGCG 440
Db 36710 TGGTGTCTACGTGGGCGCGCGCATCGGCTGTACGCGCTCGATCCCGGCTGGTCCGA 36769
QY 441 GTCCCGAGAGATCGACGCTCATGGCGGGGCTGGAGAGATCCAGCTTCGAGCGGCG 500
Db 36770 GGGCTCGAGGAGCTATCTGCTGACCGCGAGCGCGCGGCGGCGGCGGCGGCGG 36823
QY 501 AATCTCGATTCCTCGGCGCTCGAGGCGCGTGTGTCGCGGTGATATCGGCTATTCGTC 560
Db 36824 CATCTCTACGCGCTCGGCTCGAGAGACGCTTCATGACGGTGGAGACGCGCTGCTCTC 36883
QY 561 CTGCGTGTGCGCGCTTCATCTGAGCTGTGACAGCTTGGCTCGGGGAAATGCTCCAGCG 620
Db 36884 CTGCGTGTGCGCGCTTCATCTGAGCTGTGACAGCTTGGCTCGGGGAAATGCTCCAGCG 36943
QY 621 CTTGCTGTGTTGGGATTCGCTGATGTTGTCGCGAGACCTTCGTGTGCTTCCAGAGC 680
Db 36944 GCTGGCGGGGCGGGTGGCGGTATGCGGATCCGCGGGGTTCTGTGAGATTTCTCCGGA 37003
QY 681 CGGGGCGCTGGCGAGGAGCGTTCGCTGCAAGGCAATTTTCGCGAGAGCGCGATGGTTCCG 740
Db 37004 GAAAGGCGCTGGCGAGGAGCGCTGCAAGGCGTTCTTCGCGCGCGCGAGACGCGAG 37063
QY 741 AGAGGCGAAGGAGTGGCGCTGTGTGTTCTTCAAGCGGCTCAATGAGCGCGCGAGACG 800
Db 37064 CTGGGCGAGGCGGTGGCGGTGTGCTCTGAGAGCGGCTGTGAGACGCGCGCGCGG 37123

QY 801 CGATGATATTTGGCGGATTCGAGATCCGCGATCATCGACGCGTGGAGCGCG 860
Db 37124 GCAACGGTCTCTCGGCTGTACCGGACCGCGGTCAACGAGAGGTTGCTTCAACGG 37183
QY 861 TCTGACCTGCGGACGAGGAGCTCCAGAAATGCTGTGAAACGCGGCTTGGCGAGCG 920
Db 37184 GCTACCGCGCGCAACGCGCCAGCCAGCAACGCGTATCGCCGAGGGGCTTCGCGAGCG 37243
QY 921 AGGCTGCGCGCGCTTCTGTTGGTATGTGAGGACACGCGACGCGGACGAGCTTGG 980
Db 37244 CGGGCTGTCCCGGAGACGTGAGCGCGGTGAGCGGACGCGACCGGACCGGCTTCGG 37303
QY 981 TGACCCCATGCAAAATCCAGCTGTAATGCGGTATACGCGCTCGGGCGAGATGTCCAC 1040
Db 37304 GACCCCATGCAAGGCGGGGCGCTGTCTCGCGCTCCGACGAAACGTTCCGCGACCA 37363
QY 1041 GCCGCTGTGATTCGGTGTGAGAGACCACTTGGCATCTGATATGCGTGGGAT 1100
Db 37364 CCGCTGTGGCTCGGCTGTGAGAGTCCATATCGGCGATGCCAGGCGCGCGCGGTGT 37423
QY 1101 CACTGGGCTGCTGAAGTGTCTTGTCTTCCCTTACAGCGGCGAGATTCTTGGCACCTCA 1160
Db 37424 CGGCGGCTCATCAAGATGCTTCAGCGCTCGGACGCGCTGTGCCCCGACCTTCA 37483
QY 1161 CGCGAGGCGGTGAACCCCGGATCTCATGGGGTATCTTGCGCTGACCGTACGCGCG 1220
Db 37484 CGCGAGCGGCGGACCCCGCATGCGACTGAGCTCGGCGGAGTACGCTGTCACTTC 37543
QY 1221 CGGACACCTGTGGCGGACTGAAATACGCGCGACGCGGCGGGTGAAGCTGTTCGCGAT 1280
Db 37544 CGAGGTGCTGTGGAGCGGAGCGGCGCGGCGCGCGGAGTCCGCTTCGCGGT 37603
QY 1281 GAGCGGAGCAACGCGCATGTGTGTGGAAGAGCGCGCGGCGGACGTCAACCGCC 1340
Db 37604 CGGCGGACCAATGCGCATGTGTGTGGAAGAGCAACCGCGCGCGCGGAGAC 37663
QY 1341 GGGCGCGGAGGAC-----GGCAGAGCTGCT 1367
Db 37664 GGGCGGAGGACCTCGGCGGCTCCCGCGCGCGAGAGGGCGGAGGGCTTCGCGT 37723
QY 1368 GGTGTGTGCGCAAGAGACCGGCTGAGCTTGAATGCAAGCGGCGGCTGCGGACCA 1427
Db 37724 GGTGTGTGCGCAAGAGACCGGCTGAGCTTGAATGCAAGCGGCGGCTGCGGACCA 37783
QY 1428 TCTGAGACCTACCTTGTGCAAGTGTGTGGAGAGTGTGCGCTTCACTGTGGAGAGCG 1487
Db 37784 CTTCTCCGCAACCCCGGGGCGCGCGGTGAGACATGCTTCTTCGCGCGCGCG 37843
QY 1488 CAGCGCATGAGACACCGGCTCGCGTGGCGGACGTCGAGAGAGGAGGCTGCGGAGCG 1547
Db 37844 CGAGCTTTTACACCGGCGCGGTGTGATGCGCTCGGACGGGCGCAACTTCGCGCG 37903
QY 1548 CTTGACGCTCGGCGGAGAGACAGCTGCGCGCGGTGCGGCACTATCGCGATTTC 1607
Db 37904 CTTGACGCTGTTGGCGGAGAGACAGCTGCGCGCGGTGCGGCACTATCGCGATTTC 37960
QY 1608 CTCACGCGGCAAGCTCGCTTCTTCTTCAACCGGACAGGGGCGGACGCTGGGCGAG 1667
Db 37961 CCGGACCGGCAAGATGCGCTTCTTCTTCAACCGGAGGAGGACGAGCGCGCGGAGTGG 38020
QY 1668 CCGTGGGTGTATGATATGATGCTCGCGGTTCCGAGGCGCTTCACTGTGAGGCT 1727
Db 38021 CCAGCACTTCAATGCGCGCATACCTTCTTTCGCGCTCGGCTTCAAGAGTGAAGGACG 38080
QY 1728 GTTCAACGAGAGCTCGACCGGCGCTTCGCGAGGTATGTTGGCGGAAACGCGCGAGCT 1787
Db 38081 TCTGACCGCGCTGCTCGGCGCGCTGCGGCGCTGTGAGCGCGGACCGCGGCTGCGC 38140
QY 1788 CGAGCGCGCGCTGCTCGACGAGACAGCTTCAACCGCGCGGCTGTTCATCTTGAATA 1847
Db 38141 CGAAGCGGCACTTCTTGAACCGGACGAGTACACCGCGCGGCTTTCGCGCTGAGAGT 38200
QY 1848 TGGCTGCGCGCGCTGTGTGGCGGTGTGAGCGGAGTTGTGCTCGCGCATAG 1907

Db	38201	GGCGCTCCACCGGTGCTGGAGCACTGGGGGAAATGCCGCCCGACCTCTCTCTGGGGGACATCC	38260
OY	1908	CATCGGTGAAGCTGTGTGCTGCTGCTGCGTGGCGCGGTGTTCTCGCTTGAAGACGCGGTGTT	1967
Db	38261	GGTGGGGGAACTGGCGGCGCCGCCACATCGTCGGGGGTGTGCTGATCTTGAGAGACGCTTGGCG	38320
OY	1968	CTTGTGGCTACGCGCGCGGGGCGCTGAATGCGAGAGGCGCTGGCGCGCGCGCGCGCGATGTGTC	2027
Db	38321	GCTGTGTGGCCGCCCGCGCGGACAGCTGAATGACAGCGCTGTGCGCCCGCGCGCGCGATGTCTTC	38380
OY	2028	GATCGAGCGCCCGGAGGCGCGATGTGCTGTGCGTGTGGCGCGACGACGACGTCGTGTC	2087
Db	38381	CGTCCGCGCGCGCGCGAGAGGACGAGGTCCGCGCACTGTGGCGCGCGCGCGAGAGACGCGCTTG	38440
OY	2088	GATCGCGCGCGGTCAACGCTCCGAGACAGAGTGTGATATGCGGGGCGCGGGCGAACCGGTGCA	2147
Db	38441	CGTGGCGCGGTGAAACGCGCCCGCGTGTGTGTGATCTTCGAGGCGGAGAGAGCGGTGTC	38500
OY	2148	TGCGATGCGCGCGCGCGATGCGCGCGCGCGCGCGGAGCGGAAACCAAGCGCTCCACGTCTCGCA	2207
Db	38501	CGAGCGCGCGCGCGAGCTCGCCCGGAGGAGCGCGCGCGCGCGCGCGCGCGCTCGCGTCCGCGA	38560
OY	2208	TGCGTTTCCACTCACCGCTCAATGCGCCCGGATGCTGAGAGCGCTTTCGGCGGTGTGGCGGATGC	2267
Db	38561	CGCGTTTCCACTCACCGCTTGAATGAGCGGACATGCTGCGCGGATTCGCGGAGAGGTGCGCCGCGG	38620
OY	2268	GGTAGGCTAACCGCGCGCGCGCTGATATGATCTGTGTCAGAGCAATCTGAGGCGGAGG---GTTTG	2324
Db	38621	CTGTGCGTTACCGGGAAACCGGAGCTGAACGATGCTGTCTCAGCGTCAACGGGCGCGCGCGCG	38680
OY	2325	CACAGACGAGGTGAGCTGCGCGGAGCTAATTGGGTGCGCGCACGCGCGAGAGGTGTGCGCTT	2384
Db	38681	CCCCGGTGAATCTCACCGGCGCCCGCATCTATGAGTGTGCGCGACAGGTCCGTGAGACCGCGTGCTT	38740
OY	2385	CGCGGATGAGATGAAAGCGCGCTGACACGCGCGCGGTGTGGGCGACCTTGATGAGGTGCGTTC	2444
Db	38741	CGCGGAGCGCGGTCCGCGACGCGACACCGCGCTTGAGGCCCGGACCTTCTTGAGAGCCGAGCC	38800
OY	2445	GAATTCAGACGCTGTCTGCGGCTGTGTGCTGTGCTGCTGCATGTCCGAGACCGCGCGCGCTGTCT	2504
Db	38801	GGAACGCGGTCTGTGTGCGGCGAATGGAGAGAGTGTCTGTGAGAGAGACACACGTTGGCCCTGCT	38860
OY	2505	CGCATCTGTGCGCGCTGTGGGCGTGAACGAGCGCGGACCGTGTCTGTGAGGCGCTGCGCGAGCTT	2564
Db	38861	---GCCGAGCATTCACAAGGCCCGGACACCGGCGCGCACAGGTGCGCGCGCTCTCCGCGCGCGCT	38917
OY	2565	CTGGGCGGTGCGATGTGGCTGTGTCCTGTGGGCGCGGCTCTTCCCTTCAGGGGAGGCGGCGGT	2624
Db	38918	GCGGGCGCGCGCGCGCGCGCTGACACGCGCGCGGGCGCGCGCGGTGTGAGTCT--GGGCGGATGAC	38975
OY	2625	GCGCGTCCCACTGATCCCTTTGGCAGCGCGAGCGCTACTGTGATGACACGAAAGCGGACGA	2684
Db	38976	GCCACACGCGCCCGGAGGAGCGCGCGCGCGCGCGCGGTGCGAACTGCGCTCCACCTTCCGCGCAC	39035
OY	2685	CGCGGCGCGGTGCGACACGCGCGGTGTCCGCGGAGGGGTCAACGAGAGGTGAGAGAGGGGGG	2744
Db	39036	CGCGCGCTACTGTGCTGCGCCCGGGCGCGCGCGCGCGGACACGACGACTGTGATATCCGGAATC	39095
OY	2745	CGCGGTGCGCGCGCGCGACACCGCGCGACGCGCTCGGCTTGCACATCCGCGCGCGCGAGACGG	2804
Db	39096	GCGTGTGGACGCGCTGCGCGGTGTGACACCGCGGGGGGCCGAGACGCGCGCGCGCTGTGCTGTG	39155
OY	2805	ACGCGCGGAGAAAGTTCAGAGCCCGCGCGACACGTTCCGTTCCGGC	2848
Db	39156	ATTCACCCCGACAGCCCGCGCTGTGCGGAGCTGTGCGGACACGC	39199

RESULT 12	
AAV21187	
ID	AAV21187 standard; DNA; 53789 BP
XX	
AC	AAV21187;

XX	24-JUL-1998	(first entry)	
DT			
XX			
DE	Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.		
XX			
KM	Amycolatopsis mediterranei rifamycin synthesis; gene cluster;		
KM	polyketide synthase; actinomycete; ansamycin; ds.		
XX			
OS	Amycolatopsis mediterranei.		
XX			
FX	Key	Location/Qualifiers	
FX	CDS	1825..15543	
FT		/*tag= a	
FT		/label= ORF_A	
FT		/product= "polyketide synthase"	
FT	CDS	15550..30759	
FT		/*tag= b	
FT		/label= ORF_B	
FT		/product= "polyketide synthase"	
FT	CDS	30895..36060	
FT		/*tag= c	
FT		/label= ORF_C	
FT		/product= "polyketide synthase"	
FT	CDS	36259..41325	
FT		/*tag= d	
FT		/label= ORF_D	
FT		/product= "polyketide synthase"	
FT	CDS	41373..51614	
FT		/*tag= e	
FT		/label= ORF_E	
FT		/product= "polyketide synthase"	
FT	CDS	51713..5293	
FT		/*tag= f	
FT		/label= ORF_F	
FT		/product= "polyketide synthase"	
XX			
PN	NO9807866-A1.		
XX			
PD	26-FEB-1998.		
XX			
PF	18-AUG-1997;	97WO-EP04495.	
XX			
PR	20-AUG-1996;	96EP-0810551.	
XX			
PA	(NOVS) NOVARTIS AG.		
XX			
PI	Engel N, Schupp T, Toupet C;		
XX			
DR	WP1, 1998-169172/15.		
DR	P-PSDB; AAM52845, AAM52846, AAM52847, AAM52848, AAM52849, AAM52850.		
XX			
PT	Amycolatopsis mediterranei rifamycin synthesis gene cluster - used		
XX	to produce rifamycin and rifamycin analogues		
PS	Claim 4; Page 53-102; 205dp; English.		
XX			
CC	The present sequence represents a Amycolatopsis mediterranei rifamycin		
CC	synthesis gene cluster DNA fragment from the present invention. The		
CC	DNA fragment comprises a DNA region involved directly or indirectly		
CC	in the gene cluster responsible for rifamycin synthesis, including		
CC	the adjacent DNA regions to the right and left which, by reason of		
CC	their function in connection with rifamycin biosynthesis, qualify		
CC	as constituents of this rifamycin gene cluster, and functional		
CC	fragments, derivatives or constituents of these. The Amycolatopsis		
CC	mediterranei rifamycin synthesis gene cluster DNA fragment can be used		
CC	for producing rifamycin, rifamycin analogues or precursors. It can also		
CC	be used for inactivating or modifying genes involved in ansamycin or		
CC	rifamycin biosynthesis. The DNA can be used for constructing mutant		
CC	actinomycetes strains from which the natural rifamycin or ansamycin		
CC	biosynthesis gene cluster has been partly or completely deleted. The		
CC	DNA fragment can be used for assembling a library of polyketide		
CC	synthases, which can be used for assembling a library of polyketides.		
CC	A hybridisation probe of the invention can be used for identifying DNA		
CC			

CC Fragments involved in the biosynthesis of anabamycin.

XX Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

Query Match 17.5%; Score 745.8; DB 19; Length 53789;
Best Local Similarity 56.3%; Pred. No. 8.3e-111;
Matches 1674; Conservative 0; Mismatches 1232; Indels 69; Gaps 12;

QY 28 GCCGAGATCGGATTCGATCGTCGAGCGAGTTGCCGTCGCCCGGCGGTATCGAT 87
DB 41457 GCCGAGCGAGCTCGCGCTCGTCGGGATGACCTGCGGGTCCCGGGTGGGTGTCCTGC 41516
QY 88 CTGACCGGCTTCTGACCGCTCTCGAGGGCTGCGCGAACCCTGCGGGGAACTCCCGCG 147
DB 41517 CCGAGACCTGTGGACCTGTGGCCGGCGGGGTGACGACCTTTCGACTTCCCGAC 41576
QY 148 GAAC---GCTGGGATGACAGCAGTGTGATTGATCCGACCCCGATGCCCGGGGAAAGC 204
DB 41577 GACCGGGGCTGGGAGCTGAGCGGCTGTTCGACCCGACCCGACCCCGGGAGCTGC 41636
QY 205 CCGGTTACCGCGCATCTTCTCTGACGACGTAAGCTGCTTGCAGCGCTCTTTCGAGC 264
DB 41637 TACACAGCCAGGCGGCTTCTGTGCGTGGCCGCGGCTGTTCGACGCGGGCTGTTCGCG 41696
QY 265 ATCTGCGCTTCGAGAGCGCTGCGGATGACCTTCGACATCGACTCTTGTGAGAGGTTCG 324
DB 41697 ATCTGCGCTTCGAGAGCGCTTCGATGACCCGACGACGAGGCTGTGCTGAGAGCTGC 41756
QY 325 TGGGAGGCGCTGAGAGACCGCGATCGCTCCATCGCGCTGTCGGTTCGAGAAACGGGA 384
DB 41757 TGGGAGGCGCTTCGAGAGACCGCGGCTGCACCGCTTTCGTAAGGCGACGACCGTCCG 41816
QY 385 GTGTTTCATCGGATCGGCGCTTCGCAATGAGCCGCGCTCCGCAACGCGCGCTCC 444
DB 41817 GTGTTCTCGCGGCTTTCACCAAGGCTACGCGCGCGCGG-----CGATCACG 41864
QY 445 GGAAGATGAGAGCTCATGCGGCGGCTGGGAGAGATGCCAGCGTGGAGCGGGCGGAATC 504
DB 41865 CCGAGACTTCGAGCGCTTCGCGGACATCGGGGCGGCTGCAGAGCTGCGTGGGCGGGTG 41924
QY 505 TCGTATGCGCTCGGCGCTGCGAGGCGCTGTGTCGCGGATGATACGCGCTATTCGTCCTG 564
DB 41925 TCTGATGCTTCGCGGCTGCAAGGACCGGGGCTCAACATGACACCGCGCTGTGTCGTCG 41984
QY 565 CTGCTGCGCTTCATCTGCTGCTGACAGCTTGCCTGCGCGGAAATGCTCCACGCGCTG 624
DB 41985 CTGCTGCGCTTCATCTGCTGCTGACAGCTTGCCTGCGCGGAAATGCTCCACGCGCTG 42044
QY 625 GCTGGTGGGGAATCGCTGATGTTGTGTCGCGAGACCTGCTGCTGCTGCGAAGACCGG 684
DB 42045 GCGGCGGGGCGACGCTGATGTCGAGCGCGGACCTTGTGCGCTTCTGCGCGAGCGG 42104
QY 685 GCGCTGCGAGGAGCTGCTGCAAGGCAATTTTCGCGGAGGCGCGATGGTTCGACGCA 744
DB 42105 GTGCTGCTGCGAGCGCGGCTGCAAGGCTTCTCTCGACCGCGGACCGGCTGCG 42164
QY 745 GCGGAGGCTGCGCGCTGCTGCTCAAGCGGCTCACTGAGAGCCCGCGGACGCGCAT 804
DB 42165 GCGGAGGCTGCGCGGCTGCTGCTCAAGCGGCTTTCGCTGCGCGGAGAGCGCGGCTC 42224
QY 805 CGGATTTGCGCGGATTCGAGGATCGGATCAATCAGACGCTGCGAGCGCGCTGCG 864
DB 42225 CGGATTTCTGCGCGCTGCTGCGGCGAGCGGCTTCAACAGATGCGCTTCCACGCGCTG 42284
QY 865 ACCGTCGCAAGCGGAGCTCCAGAAATCTGCTGAAACCGGCGCTTCGCGAGCGAGCG 924
DB 42285 ACCGTCGCAAGCGGAGCTTCGAGAGCGGCTGATCCGAGGCGCTTCGCGGCGCGG 42344
QY 925 TCGCGCGCTTCTGCTGCTGATATGTCAGAGGACACGCGGACGAGCGCTTGTGTCAC 984
DB 42345 CTGCTGCGCTTCGATGCTGCTGAGAGCTGTCGAGCGGACCGCGGCTGCGGCGAC 42404
QY 985 CCGATGGAATCCAGCTCGATGAGGATACGGCTGCGGGGAGAGATGTCGCGACGCGG 1044

DB 42405 CCGATGGAAGCGGAGCGCTCTGCGACCTACCGGCCAGGCGCTGA-----GCGGCGG 42458
QY 1045 CTGCTGATCGGCTCGTGAAGACCAACTTGGCCATCTGATGATGCTGCGGAGTCACT 1104
DB 42459 CTGCTGCTGCGGCTCGTCAATCGAATTGCGGACACAGGAGCGCGCGCGGGTGC 42518
QY 1105 GGGCTGCAAGGCTGCTTGTCTCTTCAGACGCGGACAGATTCCTGCGCACTTCCAGCG 1164
DB 42519 GCGGATCAAGATGCTCAGAGCCCTGCGGACGCGGCCAAGCGCGGACCTTCACGCTG 42578
QY 1165 CAGGCGTGAACCCCGGATCTCAAGGAGTATCTTCGAGTCAACCGTACCGCGCGCG 1224
DB 42579 GCCAGCCGACCGGAGAGTGAATGCTGCGCGCGCTGCGGAGTGAATCTGACGAGCG 42638
QY 1225 AACCGTGGCCGGACTGGAATATCCGCGAGCGGCGGGGATGAGCTGTTTCGATGAC 1284
DB 42639 CCGGATGCGCCCGCGGATGAGCTGCGCGCGCGGCGGGGATGTCGCGTTCGGATCAGC 42698
QY 1285 GGGACCAACGCGCACGCTGCTGGAAGAGGCGCGCGCGGAGAGCTGACACCGCGCG 1344
DB 42699 GGGACCAACGCGCACCTGATCTGAGAGAGCGCGCGCGGAGAGCGCGGAGAA 42758
QY 1345 CCGGAGCGACCGG---CAGAGCTGCTGCTGCTGCGCAAGAGACCGGCTCAGCCCTGAT 1401
DB 42759 CCGGATTCGAAGGAGCGGCTGCTGCTGCTGCTGCGGAGACCCCAATCTTTGCGG 42818
QY 1402 GCACAGCGGCGCGGCTGCGGACCATCTGAGACCTAACCTTCGAGTGTTCGAGCAT 1461
DB 42819 GCTCAGGCGCGGCTGCGGAGGATCTTGG---GCTCGGATGATGTCCTCCGGGCGCG 42875
QY 1462 GTGCGCTTCACTTCGCGGACGACCGCGAGCGGATGAGACACCGGCTGCGGTCGCG 1521
DB 42876 CTGCGAGCGGCTGCTGCTGCGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 42935
QY 1522 ACGTGAAGAGGAGGCTGCGGAGCGGCTGAGCGCTGCGGCGGAGGACAGACGTCGCC 1581
DB 42936 GGAAGGACGAGAGAGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 42995
QY 1582 GATGCGGCTGCGGATGATGCGGATCTCTACGCGGAGAGCTGCGCTTCTTCAACGGA 1641
DB 42996 GCGGCTGAGACCGGCTGCGGCA-----GACGCGCAAGGCTGCTACGCTTCCCGG 43049
QY 1642 CAGGAGGCGGAGCGCTTGGGATGAGGCGGCTGATGATGATGATGATGATGATGATG 1701
DB 43050 CAGGATTCGAGCGGCTGCGGATGAGGCTGAGCTTACGACCGGATCCGCTGCTGCG 43109
QY 1702 GAGGCTTCGACCTGCTGCTGAGCTGTTCAACGAGAGCTGACCGGCGCTGCGGAG 1761
DB 43110 ACGGCTTCGAGGAGCTGCGGAGCTGAGACGCTGCTGCTGCTGCTGCTGCTGCTG 43169
QY 1762 GTGATGAGGCGGACCGGCGGAG-----GCTGACGCGCGGCTGCTGACCAAGCA 1812
DB 43170 CCGGCTGCGGAGCTGCTGCTGCGGAGGCTGCGGCGGAGGCTGCTGCTGCTGCTGCTG 43229
QY 1813 GCTTCAACCGGCGGCTGCTTCACTTGAATATGAGCTGCGGCTGCTGCGGAGCTG 1872
DB 43230 GCTTCAACCGGCGGCTGCTTCACTTGAATATGAGCTGCGGCTGCTGCGGAGCTG 43289
QY 1873 TGGGATGAGAGCGGAGTGTGCTGCGGCGGATGAGCTGCTGAGCTGCTGCTGCTGCTG 1932
DB 43280 TGGGATGAGAGCGGAGTGTGCTGCGGCGGAGTGTGCTGCGGCGGAGTGTGCTGCTG 43349
QY 1933 GTGCGGCGGCTGCTTCTGCTGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1992
DB 43350 GCGCGGCGGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43409
QY 1993 ATGCAAGGCTGCGGCGGCGGCGGAGTGTGATGAGGCGCGGAGCGGAGCGGATG 2052
DB 43410 ATGCAAGGCTGCGGCGGCGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43469
QY 2053 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2112

Db 43470 GC-----CGACTGCTCGGCGAGCGGCGTGAATCCGCCCGCTTCACAGCCCTTCG 43520
 Qy 2113 CAGGTGTCATCGCGGCGCGCGGCAACCCGTGATGCGATCGCGGCGCGATGCGCGC 2172
 Db 43521 GCGGTATCTCTTTCGCGGGAACGCGGACGCGGTCTGTCGCGCGCGCGCGCGATCGCGAG 43580
 Qy 2173 CGCGGCGCGGACCAAGGCGCTCAAGTCTGATGCGATGCGTTCATCAACGCTCATGAGCC 2232
 Db 43581 CGCGGCGCAAGACCAAGCAAGCTCAAGTTTCGACGCGTTTCATCTCCGCGCGATGAGCG 43640
 Qy 2233 CCGATGCTGAGAGGCTTCGCGGCGGTGCGCCGAGTCCGATGAGCTACCGCGCGCTCATC 2292
 Db 43641 CCGATGCTGAGAGGCTTCGCGGCGGTGCGCCGAGTCCGATGAGCTACCGCGCGCTCATC 43700
 Qy 2293 GTCTGATGACAGATCTGA---GCGGAAAGGCTTTCACAGAGAGTGAAGTCCGCGGCG 2349
 Db 43701 CCGGTGCTTCACAGTGAACGCGCGCGGTTCGCGCGAGCCGCGCAACTGACCGAGCCGCGC 43760
 Qy 2350 TATTGGGTGCGCACGCGCGAGAGGTGCGCTTCGCGATGAGAGTGAAGGCGCTGAC 2409
 Db 43761 TACTGGGCGGACAGTGCAGCGCGCGGTTCGCGCGAGGCGTTCGCGCGCGACG 43820
 Qy 2410 GCGGCGGCTGCGGCGACCTTGTGAGAGTCTGATCCGAATGACGCTGCTGCGCTGTG 2469
 Db 43821 GAGTCCGCGCGCTGCTGCTGCTGAGAGTCTGCGCGCGCGCGCGCGCTGACCGCTCTGT- 43879
 Qy 2470 CTTGCTGATGCGCGAGCGCGCGCGCGCTGCTGCGATGCTGCGCGCTGCGCGCGTGA 2529
 Db 43880 -----CGAGGAGAGCGCGCGAGTCACTGCGCTGCGCGCGCGCGCGCGCGCGCG 43931
 Qy 2530 GAGCGGCGAGCGGTCTGAGAGCGCTGCGCGCGGTCTGCGCGCGTGTGCGCTGTCC 2589
 Db 43932 GAGGTCAACGCGCGTGAATCAACGCGCGGTGCGCGAGCTGTTGCGCGCGGAGTTGCGGTGAT 43991
 Qy 2590 TGGGCGCGCTCTTCCCTCAAGGCGCGCGCGGTGCGCGCTGCGCGCGTGTGCGGTG 2649
 Db 43992 TGGCGCGCGCTGCTGCGCGCGCGGTGCGCGCGGTGCGCGCGTGTGCGCGCGTGTGAC 44051
 Qy 2650 CGCGAGCGCTACTGATGACACGAAAGCGAGCGCGCGCGGTGCGCGCGTGTGCT 2709
 Db 44052 CAGAGCACTATTGCTGCTGCGCGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCA 44110
 Qy 2710 CCGGAGCGCGGTCAAGAGAGTGAAGAGGCGCGCGGTGCGCGCGCGCGCGCGCGCGCG 2769
 Db 44111 GGTGCGCGCGCGACCAACCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGGTGCGCGCGCT 44170
 Qy 2770 AGCGCTGCGCTGACATCCGCGCGCGAGAGCGCGAGCGCGGAGAGTGAAGCGCGCGC 2829
 Db 44171 GGTCTTCACTCTG-----CGGCTGTCAATTGAATCGACCGCGTGTGCGCGCGCGTGC 44225
 Qy 2830 GCGGCGGCTCGCTCGCGGTGAGATGAGAGCGCGCGGTGCTGATCACTGTGCTT 2889
 Db 44226 ATCCGCGCGGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44285
 Qy 2890 CCGGTCAAGAGCGCGCGCGCGCGCGGTGCTGCGCGAGTGAAGTGAAGTGAAGCGCGCGC 2949
 Db 44286 GAGAGCGCGCGCGCGCGCGGTGCTGCGAGAACTGCTCAATCGAGGCTCGCGTGTCTCC 44345
 Qy 2950 GGGCTCAGCTTCAATGATGATTCAGCTGCGCGTGG 2984
 Db 44346 GACCAAGCGCGCGGTCCGAGTCCAGTGTCTGTGG 44380

RESULT 13

AA287283 standard; DNA, 15872 BP.

AA287283;

05-JUN-2000 (first entry)

S. venezuelae vep ORF 1, SEQ ID NO:1.

XX

KM Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KM neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KM biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KM chronic obstructive pulmonary disease; respiratory inflammation;
 KM hypercholesterolemia; crop protection agent; ds.

Streptomyces venezuelae ATCC15439.

Key Location/Qualifiers

FT CDS /tag= a
 /product= "vep ORF 1 amino acid sequence #1 (AAV77177)"

FT CDS /tag= b
 /product= "vep ORF 1 amino acid sequence #3 (AAV77199)"

FT CDS /tag= c
 /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"

FT CDS /tag= c
 /product= "vep ORF 1 amino acid sequence #2 (AAV77178)"

MO20000620-A2.

06-JAN-2000.

25-JUN-1999; 99MO-US14398.

26-JUN-1998; 98US-0105537.

(MINI) UNIV MINNESOTA.

Sherman DH, Liu H, Xue Y, Zhao L;

WPI; 2000-160679/14.

P-PSDB; AAV77177, AAV77178, AAV77199.

Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

synthesis of methymycin and pikromycin.

Example 3; Figure 23; 438bp; English.

CC The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the erylC gene cluster of *Saccharopolyspora erythraea* or
 CC *Streptomyces antibioticus*. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics,
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants. The
 CC present sequence represents a *Streptomyces venezuelae* ATCC 15439 DNA
 CC sequence, designated vep ORF 1 in the specification, which actually
 CC contains 3 open reading frames, which encode proteins AAV77177-Y77178 and
 CC AAV77199. The vep ORF 1 protein is defined in the specification as a PHA
 CC monomer synthase.

Sequence 15872 BP, 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

Query Match 17.4%; Score 742; DB 21; Length 15872;

Best Local Similarity: 55.7%; Pred. No. 3,6e-110;
Matches 1594; Conservative 0; Mismatches 1215; Indels 53; Gaps 7;

```
OY 26 CAGCCGAAGATCCGATTCGATGTCGAGCGAGTTCGCTGCTGCCCCGTCGATCG 85
DB 8490 CCGAGGACGATCCGATCCGATCGTGGGATGCGATGCGCTACTCCGGGTGGTGTGCT 8549
OY 86 ATCTAGAGGGGTTCTGAGCGCTCTCGAGAGGCTCCCGGACACCTTCGGGCGAATGCC 145
DB 8550 CGCCGAGAGACTGTGGCGGGCTGTGGCGGAGGAGGACGACGCGATCAAGCGAGTTCCCG 8609
OY 146 CCGAAC---GCTGGATGACGAGCGGTGTTGATCCGACCCCGATGCCCGGGAGAGA 202
DB 8610 TCAACGGGGGCTGGAGACCTGAGAGCCTCTACGACCCCGATCCGAGTCCGAGGACCA 8669
OY 203 CGCCGTTACGCGCGCATCTTTCTGAGCGACGTAGCCTGCTTCGACGCTCTTCTTTCG 262
DB 8670 CGTACTGCGGGAGGGCGGGTTCCTGGAAGCGCGCGTGACTTCGACGCGCTTCTTTCG 8729
OY 263 GCATTCGCTTCGCGAAGCGCTGCGGATGGAACCTTGCACTGACTTTCCTGAGGTCT 322
DB 8730 GCATTCGCTTCGCGAAGCGCTGCGGATGGAACCTTGCACTGACTTTCCTGAGGTCT 8789
OY 323 GCTGGAGAGCGCTGAGAAACGCGCGATCGCTCAATCGGCGCTCGTGGTACGAAACG 382
DB 8790 CTTGGAGAGCGCTGAGAAACGCGCGATCGACCTGCTCTGCTGCGCGGACCGCGGTG 8849
OY 383 GAGTTTCAATCGGATTCGCGCGCTCGAATATAGAGCGCGCGTGCAGACGAGCGGT 442
DB 8850 GTGTCTAGTGGGCGCGCGGACGCGCTGTGACGCTCGATCCCGGCTGTGGCCGAG 8909
OY 443 CCGGAGATGCAACCTCATGCGCGGGCTGGGAGCAATGCCAGCTTCGAGGGGCGAA 502
DB 8910 GCTCGAG---GCTATCTGTGACCGGAGCGCGGCGGATGATGTCGCGCCCA 8963
OY 503 TCTCTATGCTCTCGGAGCTGCGAGGCGGTGTGTGCGGTGATACGCGCTATTGCTCT 562
DB 8964 TCTCTAGCGCTCGGTCTCGAAGGACGCTCATACGCTGAGAGCGGCTGCTCTCTCT 9023
OY 553 CGCTGTGAGCGCTTATCTGCTGTCTGAGCTTGCCTCGGGGAAATGCTTCACGCGCC 622
DB 9024 CGCTGTGAGCGCTTATCTGCTGTCTGAGCTTGCCTCGGGGAAATGCTTCACGCGCC 9083
OY 623 TGGCTGTGGGGTATCGCTGATGTTGTGCGCGAGACCTCTGTGTGCTCTGAGACCC 682
DB 9084 TGGCGGGGGGTGCGGTGATGTCGCGATCGCGCGGCTTCTGTGAGTTCTTCGCGAGA 9143
OY 683 GGGCGCTGAGCGGACGCTCGCTCAAGGCAATTTTCGCGGAGCGCGATGAGTTTCGAC 742
DB 9144 AGGGGCTGGCGCGGACGCGCGCTGCAAGGCTTCTCGCGCGCGCGGACGCGACCGCT 9203
OY 743 GAGCGAAGGTCGCGCGCTGTGTCTCAAGCGGCTCAGTGAGCCCGCGGACGCGCG 802
DB 9204 GGGCGGAGGTCGCGCGCTGTGTCTCAAGCGGCTCAGTGAGCCCGCGGACGCGCG 9263
OY 803 ATCCGATATGCGCGGTGATTCGAGATTCGCGGATCAATCAACGCGTCAAGACGCGTC 862
DB 9264 ACAAGCTCTGCGCTGTGTCAACGCGACCGGCTCAACGAGACGATGCTTCAACGAGCC 9323
OY 863 TGAACGTCGCGGACGCGGATCCCAAGAAATGTCCTGAAAGGGGCGCTGCGGAGAG 922
DB 9324 TGAACGTCGCGGACGCGGATCCCAAGAAATGTCCTGAAAGGGGCGCTGCGGAGAG 9383
OY 923 GCTGCGCGGCTCTTCTGCTGAGTATGTCGAGGCAACGCGGACGCGGACGAGCTTGTG 982
DB 9384 GCTGCTGCTGCGGAGAGCTGAGCGCGGTCGAGGCGGACGCGGACGCGGCTGCGG 9443
OY 983 ACCCATGGAATTCAGGCTGGAATGCGGTATAGGCGCTCGGGGAGATGTCGCAAGC 1042
DB 9444 ACCCATGGAATTCAGGCTGGAATGCGGTATAGGCGCTCGGGGAGATGTCGCAAGC 9503
OY 1043 CGCTGCTATCGGCTCGGTGAGACCAACTTGGCATCTGAGTATGCGTGGGAGATCA 1102
DB 1102 CGCTGCTATCGGCTCGGTGAGACCAACTTGGCATCTGAGTATGCGTGGGAGATCA 1102
```

```
DB 9504 CGCTGTGCTGCGCTCGGTGAGTCAACATCGGCGATGCGCGGCGCGCGGTGCG 9563
OY 1103 CTGGGCTGCTGAAGGTGCTTGTTCCTTACGACGCGGACAGATTCTGCGACCTTCCAG 1162
DB 9564 GCGGCTGATCAAGATGCTCCAGGCGCTGCGGACGCGCTTGTGCGCGACCTTCCAG 9623
OY 1163 CGCAGGCGCTGAGCCCCCGGATCTCATGGGTGATCTTTCGCTGACCGTCAAGCGCGCC 1222
DB 9624 CCGAGAGCGGACCTCCGATGACGCTGAGAGCTCCGCGGAGTACGCTGCTACCTTCG 9683
OY 1223 GAGACCGTGGCGGACCTGAAATACGCGGAGCGGCGGAGTACGCTTTCGCGATGA 1282
DB 9684 AGGTGCGGTGGACGCGGACGCGGCGCGCGGCGGAGCGGAGTTCGCTTTCGCGCTG 9743
OY 1283 GCGGAGCTCAAGCGGACGCTGCTGAGAAAGGCGCGGCGGACGCTGACACCGCGG 1342
DB 9744 GCGGAGCTCAAGCGGACGCTGCTGAGAAAGGCGCGGCGGAGTTCGCTTTCGCGCTG 9803
OY 1343 CGCGGAGCGGAC-----GCGAGAGCTGCTG 1369
DB 9804 CCGGAGAGCGGCGGCGGCTGCGGCGGCGGCGGAGGAGGCGGAGGCGGCGGCGGCG 9863
OY 1370 TGTGTGCGCAAGGACCGGCTGACGCTGATGACAGCGGCGGCGGCTGCGGACCATC 1429
DB 9864 TGTGTGCGCAAGGACCGGCTGACGCTGATGACAGCGGCGGCGGCTGCGGACCATC 9923
OY 1430 TGAAGACTTACCTTTCGAGTGTGCGGCGGATGCTGAGTGTGCGGACGCGGCA 1489
DB 9924 TCTTCGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 9983
OY 1490 GCGGATGAGACCAACGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1549
DB 9984 CAGCTTTGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 10043
OY 1550 TGAAGCTGCGGCGGCGGAGGACGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 1609
DB 10044 TGAAGCTGCGGCGGCGGAGGACGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 10110
OY 1610 CAGCGGCGGAGCTGCGCTTCTTTCACCGGACGAGGCGGCGGCGGCGGCGGCGGCGG 1669
DB 10101 GAGAGCGGAGATGCGCTTCTTTCACCGGACGAGGCGGCGGCGGCGGCGGCGGCGG 10160
OY 1670 GTGGCTGTGAGTATGCTGCGGCTGCGGAGGCGGCTTTCGACTGCGTGAAGCTGT 1729
DB 10161 AGGACTGTGAGTATGCTGCGGCTGCGGAGGCGGCTTTCGACTGCGTGAAGCTGT 10220
OY 1730 TCAACGAGAGCTGAGCGGCGGCTGCGGAGGCGGCTTTCGACTGCGTGAAGCTGT 1789
DB 10221 TGAAGCTGTGAGTATGCTGCGGCTGCGGAGGCGGCTTTCGACTGCGTGAAGCTGT 10280
OY 1790 AGCGCGGCTGCTGAGCAGACGCTTTCACCGGCGGCGGCGGCTTTCGACTGCGTGA 1849
DB 10281 AAGCGGCACTTCTGAGACGAGCAGGATCAACCGGCGGCGGCTTTCGACTGCGTGA 10340
OY 1850 CGCTGCGGCGGCTGAGCGGCTGAGGAGTGAAGCGGAGTGTGCGGCGGCGGCGGCGG 1909
DB 10341 CGCTGCGGCGGCTGAGCGGCTGAGGAGTGAAGCGGAGTGTGCGGCGGCGGCGGCGG 10400
OY 1910 TCGGTAGCTGTGAGCTGCTGCGGCGGCGGCTGCTGCTGCTGAGAGCGGCTGCTTC 1969
DB 10401 TGGGCGAATCGGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10460
OY 1970 TGTGTGCTGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2029
DB 10461 TGTGTGCTGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10520
OY 2030 TCGAGCGGCGGCGGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2089
DB 10521 TCGGCGGCGGCGGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10580
OY 2090 TCGCGCGGCTCAAGCTTTCGAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2149
DB 10581 TCGCGCGGCTCAAGCTTTCGAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10640
```

QY	2150	CGAATCGGGGGGAGATGAGCGCGCGCGGGGCGGAAACAAAGGCGCTCAACGATCTGGCAATG	2209
Db	10841	AGCGGGGGGGCGAAGCTGCCCGGACGAAGCCCGGACACMAAGCGGCTTCGCGATGCGGACAG	107000
QY	2210	CGTTCCATCAACCGCTCATAGAGCCCGAATGCTGAGAGCGTTTCGGGCGATGAGCGAGTCCG	2269
Db	10701	CTTTCCTCACTCAACCCCTGATGAGACGAGATGCTCCGCGAATTCGGGAGAGTTCGCGCCCGGCC	10760
QY	2270	TGAGCTACCGGGCGGGCGSTGCATGCTGCTGTCTCAGCAATCTGAGCGGGAAAG---GCTTGA	2326
Db	10761	TGCGCTACCGGAAACGGAAGCTGACGGTCTCTTCACAGGTCACAGGGGCGAGCCCGCGCGC	10820
QY	2327	CAGACGAGGTGAGCTCGCCCGGGCTAATGAGTGCACGACGCGCAGAGGTGATGCGCTTCG	2386
Db	10821	CCGGTGAACTCACCGGCCCCGACTACTGGGTGACCCAGAGTCGGTAGAGCCGATGCGCTTCG	10880
QY	2387	CGGATGAGTGAAGAGCGCTGACACGCGCGGATGCGGGGACCTTTCCTGAGAGTTCGATCCGA	2446
Db	10881	CGGACGCGGTCCGACACGACACACCGGCTCGGAACCCGACCTTCTGAGAGACCGGCGCGG	10940
QY	2447	AATCGACGCTGTGCGGCTGTGATCTGCTGACCTGACATGCGGAGCGCCCGCGCGCGCTTCG	2506
Db	10941	ACGGCGGTGCTGTGGGGATGAGGACAGAGAGTGCCTGGAGAGACACACCGATGCCCTCTCTCC	11000
QY	2507	CATGTGTGCGCGCTGGGCGTGAACGAGCCGCGCAACCGTGCTCGAGGCGCTTCGCGGGCTCT	2566
Db	11001	---CGGGGATTCACAAACCCCGGACACCGCGCGCGACGATCCGCGGGCGCTCCCGCGCGCTCC	11057
QY	2567	GGGCGGTGCGGTGAGCTGTGCTCTCTGAGGACCGGCGCTTTCCTCCCTCAGGGGGGCGCGGGATGC	2626
Db	11058	GGGCGGCGCGCGCGCGCGCTGACGCGCGGGGCGCCCGGGTGAGCT---GGGCGGGATGACACC	11115
QY	2627	CGCTGCCCACTAACCTTGGCGAGCGCGAGCGCTACTGATTCGACACAGAAAGCCGACGAGC	2686
Db	11116	CGACGCGCTCCGAGAGGGGCGCGCGCGCGCGCTGCAACTTCCCGTTCACAGCGCTTCCGGACCG	11175
QY	2687	CGGCGGCTGAGCGACCGCGCGTGCCTCCGGAGCGGGGTACGACGACGAGCTGAGATTCACGAGGGGGGCG	2746
Db	11176	CGCGTACTGGCTGCGCCCGGAGCGCGCGCGCGGACACCGACAGCTGATTTACCGGATTCGG	11235
QY	2747	CGGTGCGCGGCGCGACCGGCGGACGAGCGCTCGCTTCGACCATTCGCGCGCGCCGAGAGCGGAC	2806
Db	11236	CTGGGACCGGCTGCGGCTGTGACCGGCGGGGCGCCGGAACGCGCGGCGCTGAGCTGTGAT	11295
QY	2807	GCCGGAGAAAGTTCAGAGCGCGCGCGCGACCGTCTCTTCCGGC	2848
Db	11296	CCACCCGACAGCCCGCGCTGCGGGGAGCTGTCCCGGCAACGC	11337

RESULT 14	
AAI99682/c	
ID	AAI99682 standard; DNA; 4411529 BP.
XX	
AC	AAI99682;
XX	
DT	15-JAN-2002 (first entry)
XX	
DE	Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX	
KW	Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome; variation; epidemiology; patient treatment; epidemic monitoring; ds
OS	Mycobacterium tuberculosis.
XX	
PN	US6294328-B1.
XX	
PD	25-SEP-2001.
XX	
PF	24-JUN-1998; 98US-0103840.
XX	
PR	24-JUN-1998; 98US-0103840.
XX	

PA (GENO-) INST GENOMIC.RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
XX 1551 and H37Rv differ --
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AI195683) and
CC H37Rv (AI195682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
at seqdata.uspto.gov/sequence.htm?DocID=6294338B1.
XX
XX Sequence 4411529 BP; 756565 A; 1449983 C; 1444602 G; 756379 T; 0 other;

[illegible]

OY	616	ACGGCCCTGGGTGGGGGATTCGCTCATGTGTGGCCAGACACCTCGTGGCTTCG	675
Db	3297129	TTGGGTTTGGCCGGTGGGGATACGATGTATGGGGCTGCGCTGATTTTGTGGGGTTTCC	3297070
OY	676	AAGACCCCGGGCGCTGGCCAGGGAGCGATTCGCTGCAGAGGCATTTTTCGCGAGGCCGATAGG	735
Db	3297069	CGGCAAGCGGGGGTTTGGCTGCGCATGGGAGATGTAAAGGGTTTGGCGCGGGCCGACGGC	3297010
OY	736	TTCCGACGAGGCCGAAAGGTGCGCGTGTCTTCAAGCGGCTCAGTGAAGCCCGCGC	795
Db	3297009	ACCGGTGTGGGCGAGGGGCGCGGGGTGGTTGTCTGAGCGGCTATCGACACGCGCGCGG	3296950
OY	796	GACGGCCGATCGGATATTGGCGGTATTCGAGAGATCCCGGATTCATTCAGACGCGTGGCAGC	855
Db	3296949	TTGGGGCATTCGGTGTCTGGCGGTGTGTGGGGGAGCGCGGTCAACAGAGATGGTGTCTTC	3296890
OY	856	AGCGGTCGACCGGTGCGGACGAGAGAGCTCCGAGAAATCGTCTGAGAAAGGGCCCTGGCG	915
Db	3296889	AATGGGCTGACCGGGCGCCCAATGGGCTCGCCGACGAGAGATTCAGATGGCGTTGGCC	3296830
OY	916	GACCGAGGCTGCGCGCGCTCTTCGATGGATTATGTCAAGGCAACGGCAAGGGCAGCAGC	975
Db	3296829	AACCCGGGTGTGTCGCGCGCAGAGGTAGACGTGTGTGAGGCAACAGGGAACGCAACAGC	3296770
OY	976	CTTGTGTACCCCATCGAAATCCAGAGCTCTGAATCGGATATACGGCTTCGGCGAGATGTC	1035
Db	3296769	TTGGGTGTATCCCATCGAAAGCCGAGGCAATGTCTGTCCACTATACGGGACAGGAGTCCAGCG	3296710
OY	1036	GCCACGCGCGCTGTGATATCGGCTCGGTGAGAACCAACCTTGGCCATCTCTAGATATGCGTCG	1095
Db	3296709	GAGCAGCGCGCTGTGGGTGGGTCTCAATCAAGTCGAACATGGGCCCAACCCAGCGCCGCGC	3296650
OY	1096	GGGATCTACTGGGCTGTGAAAGTGTCTTGTCCCTTGACGACGGGCAGATTCTTGCGCAC	1155
Db	3296649	GGGGTCGCGGGGTGATCAATGATGTGACGGCGCATGCGTCAATGAGGGTGATATGCGGCGACG	3296590
OY	1156	CTCCACGCGGACGCGGCTGAAACCCCGGATCTCATGGGGTATCTTCGGCTGACCGTACG	1215
Db	3296589	TTGATATGTGATGTAAGCTTCTCCCGCGGTGATTTGACCAAGCGGCGGCTGTCCGTTTGG	3296530
OY	1216	CGCCGCCCGGACACCGTGGCCGGACTGGAATATCGCCGCGGACGAGGGGGGTGAGCTGTTCC	1275
Db	3296529	ACCGAGCGCAGAGAGTGTCCGTTGAACGGGCGTCCGCGCGGGCGGGGTGTCTCTCTT	3296470
OY	1276	GGCATGAGCGGGGACCAACGCGCACGTGTGTCTGGAAGAGCGCGCGCGAC-----	1328
Db	3296469	GGGATTCAGCGGCAACACGCGCACTGTATCTTGGAAAGGGCCCCCGTACACGCCCCGCA	3296410
OY	1329	-----GTGCACACCGCGCGCGCGCGCGGAGACACCGGACAGTGTCTGG	1369
Db	3296409	GAGGCGCCGGTTGAGGCGAGGAAAGCAACCGGGGGCGCGCGCCGTCAATGATGTCCGTGGG	3296350
OY	1370	TGCTGTGGGCAAGGACCGCGTCAGCGCCCTGAGATGACAGGCGGCGCGGCTGCGGCAACATC	1429
Db	3296349	TGATTTCCGCGAGATCGGCTTAAGGCTTTGAACGGCGCAGCGGGGTGCACATTATGGCCACG	3296290
OY	1430	TGAGAACCTTACCTTTCGAGATGTGTGGGCGATGTGGCGTTTCAGTCTTGCGACAGACGCGCA	1489
Db	3296289	TGCAGGCGCAACCAAGGGCTGATCCGATCGATGTGTGGGCTGTCTGTGGCCAGT---CGCT	3296233
OY	1490	GCGGATGAGACACCGGCTCGCGGTGGCGGCGACAGCTGAGGGAGAGGGCTGCGGGCAGCC	1549
Db	3296232	CGGTGTTTGACACCGCAAGCGGTGTGTGTCCGCGCAAGCCGTGAGCAACTGATTGCCGGGC	3296173
OY	1550	TGGAACGTCGGGCGGAGGACAGACGTCGACCGCGGTGGCGGTGCGGAGTATGCGCGATTCCT	1609
Db	3296172	TGGCTGGGCTTCGCGCGGCGGACAGCCGGGTGCTCGGCGTGGCGTACGACAG---GGT	3296116
OY	1610	CACGCGCAAGCTCGCTTTTCTTTTCAACCGGACAGGGGGCGCAACCTGCGGACATGGAGCC	1669
Db	3296115	CGGTGGGCAACAGGTGTGTCTGTTTTCTTGGGCAAGGCGCGCAACGATCGGATGGGCC	3296056
OY	1670	GTGGGCTGTACGATGTATGTGTCCGCGTTCCGCGAGGCGGTTGACCTGTGCTGAGCGCTGT	1729

Db	3296055	CGAGATTGACGGGAGATTGCCCCGTTTGTGGCGACGGCATTCGATGCGGTGGCCGACGAGT	32959986
Qy	1730	TCAACAGAGACTCGACCAGCCGCTCCGCGAGGTGATGTGGCCGAAACCGGACAGCTCG	1789
Db	3295995	TGACACGGGCATCTGCGGTGGCCGCTGCGCGACGTTATTGGG-----GTGCCG	32959488
Qy	1790	ACGCCGCGCTGTGACACGACACACCTTCACCCGACCGCGCGCTGTTCACTTCGAAATATG	1849
Db	3295947	ATGCGGATTTGCTTGAACGACCCCAATTTGTCTCACCCCGCGTGTGTGCGGTGGAGGTGG	32958888
Qy	1850	CGCTCGCCGCGCTGTGGCGGTCGTGGGGTGTAGAGCCGGAAGTTGGTGCAGCGGCATATGCA	1909
Db	3295887	CATGCTTGGCGGTGTTTGCGGGAATGGGGTGTGCTTCGGACTTGATATGGGTCACTCCG	32958288
Qy	1910	TGCGTAGACTGTGGCTGCTGCGGTGGCGGCGGTGTTCCTGCTTGAAGACGCGGTGTTCC	1969
Db	3295827	TTGGAGAGCTGGCGGCGCGGACGCGCGCGGTGTGTTGACGTTGGCGGACGCGCGCATGCG	32957688
Qy	1970	TGTGTGCTGCGCGCGCGGCGCTTATGCAAGCGCTGTCCCGCGCGCGGGCGCATGTTGTGCA	2029
Db	3295767	TGTGTGTGGCGCGCGCGCGCGGCTTGTATGACAGCGCTGCGCGACGCGGCTGTGCAATGTTGCGG	32957088
Qy	2030	TGAGAGCGCGGAGGCGGATGTGGCTGTGCGGTGGCGCGCGCACGACGCTGCGTGTGCA	2089
Db	3295707	TGCGTCCGATGAGAGACGAGGT-----GAGACCGCTGTCTGGGTGAGGGTGTGGGGA	32956578
Qy	2090	TGCGCGCGGTTCAGCGCTTCGCGACCGAGTGTGATCGCGGCGCGCGGCAACCCGTGCATG	2149
Db	3295656	TGCGTGGGATTCAGCGCGCGCGGATCGGAGGTGATCTCCGGTGGCGGACCGCGGCAATG	32955978
Qy	2150	CGATGTCGGCGCGCATGCGCGCGCGCGCGCGCGCGGACGACGAGCGCTCAAGTCTGCGATG	2209
Db	3295596	CGATTTCGCGATCGGTTGCGCGCGCGGAGGTGCGCGGCTGACCACTGTGCGGTCTGCGATG	32955378
Qy	2210	GGTTCACCTACCGGCTATGGCGCCCCGATGTGTGAGAGGGGTTGCGGCGGTGGCGGAGTCCG	2269
Db	3295536	GGTTTCATTCCCGCTTGTATGGAGCCGAGAGCTCGAGAGTTGCGCGGTGTGCGGCGCGCGG	32954777
Qy	2270	TGAGCTACCGGCGCGCGCTGATTCGTCTGTGTCAAGCAATCTTGAGCGGAGAGGCTTTCACAG	2329
Db	3295476	TGCAGGACACGAGACCCCGAGTTGGGCTGGGTGCGGTGTGAGAACGTGACGGGAGATGTGCGGCG	32954177
Qy	2330	ACGAGGTGAGCTCGCCCGGCGCTATTGGGTGCGGCCACGCGCGGAGAGGTGTGCTTCCGCG	2389
Db	3295416	CTGATTTTCGGGTTCGCGCGGACGATGAGGTGAGCAACGTTTCGTGCGCGCGGTGCGCTTCCGCG	32953577
Qy	2390	ATGAGGTGAAGGCGCTGACGCGCGCGGTGGGACCTTCGTGAGAGTCCGATCGGAAT	2449
Db	3295356	ACA GTGCGGCTCATTTTGCAGACCTTTGGGCGACCCACTTCATATGAGGCGGAGCCCGGGA	32952978
Qy	2450	CGAGCTGTGCTGCGCGCTGTGCTGCTGCTGATGTCGAGCGGACGCGCGGCGCTGTCTGCAT	2509
Db	3295256	GTGGTTTGACCTGGGCTGCATTCGAGGACGTCCTTTGGCCCCGCGCTGAGGCGATGGTGTGTGCA	32952377
Qy	2510	CGTGGCGCGCTGGGCGGTGACGAGCGCGGCGACCGTGTCTGAGGCGCTGCGCGGCTTCTGGG	2569
Db	3295236	TGCTGTGGCAAAAGACCGGCGCGGAGCTGTGCGGTCTGCGGTGTGCGCGGTCTGAGGTGTTC	32951777
Qy	2570	CGGTGCGGTGTGTGTCTCCCTGGGCGCGGCTTTCGCC---CTACGAGGAGGCGGCGGGGTGC	2626
Db	3295176	CCACCGGCTGTGCGGTGACGTGTGCGGCGGTTCGCGGCTCGGGTGAGAGGCGGCGGTGC	32951177
Qy	2627	CGGTGCGCATCATCTTGTGAGGCGGACGCTATCTGATATGACACGAAAGCCGACGACG	2686
Db	3295116	AGCTGCCACGATATGCTTTTACGAGACGCGCGGTTTGTGGAGACCCCGGCGCGGATGGCG	32950577
Qy	2687	CGGCGCGGTGCGACCGCGGTGCTTCGCGGAGCGGCTGACGACGAGTGCAGGAGGCGGCGG	2746
Db	3295056	CGGCGGATGCGCGCGGCTGTGGGTCTGGGCGCGGACCGGACATGCTTGTGTGGGTGCGGTGG	32949977
Qy	2747	CGGTGCGCGGCGG 2758	

Oy 1223 GGACACCGTGGCCCGGACTGGAATACGCCGCGACCGGGGGGTGAGCTGTTCCGCAAGA 1282
 Db 9684 AGGTGCTCTGGAGCGGACCGGCGCGGCGCGGAGCGAGGTGTCCTTCCTCCGCGTCG 9743
 Oy 1283 GCGGAGCAGGCGCGGACTGGGTGTGGAAGAGCGCGCGCGCGGAGCTGCAACCGCGCG 1342
 Db 9744 GCGGACCAATGTCCATGTCTCTCGAAGAGGACCGCGCGCGCGCGCGCGCGCGCGCG 9803
 Oy 1343 GCGCGGAGCGAC-----GCGAGAGCTGCTGG 1369
 Db 9804 CCGGGGAGGCGCGCGCGCGCTCCCGCGCGCGGAGAGGCGCGCGCGCGCGCGCGCTGG 9863
 Oy 1370 TGTCTGTCCGCAAGACCGCGCTCAAGCTTGAATGCAAGCGCGCGCGCGCGCGCGCGCTC 1429
 Db 9864 TGTCTGTCCGCAAGCGCGCGCGCGCTGCGCGCTCCAGCGCGCGCGCGCGCGCGCGCG 9923
 Oy 1430 TGGAGACCTACCTTCGAGTGTCTGGGCGATGTGGCGCTTCAAGTCTGGCGAGCGCGCA 1489
 Db 9924 TCTCCCGGACCG 9983
 Oy 1490 GCGCGATGAGACACCGGCTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1549
 Db 9984 CAGCTTTGACACCG 10043
 Oy 1550 TGGAGCGTGGCGCGGAGGAGACAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1609
 Db 10044 TGGAGCGTGGCGCGGAGGAGACAGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 10100
 Oy 1610 CACGCGGAGCTCGCTTTCTCTTCAACCGGACAGGGGCGCGGAGCGCTGGGCAATGAGCC 1669
 Db 10101 GGGAGCGGAGGATGCGCTTCTCTTCAACCGGAGGAGGAGCGCGCGCGCGCGCGCGCG 10160
 Oy 1670 GTGGCTGTAGAGTATGTGTGCGCTTCCGCGAGGCGCTTGAAGCTGTGAGCGTGT 1729
 Db 10161 ACAGCTGCAATGCCCGCATACCTTCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 10220
 Oy 1730 TCAACGAGAGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1789
 Db 10221 TCGACCGCGCTGTGGCG 10280
 Oy 1790 ACGCGCGCTGTGCAACGAGAGCGCTTCAACCGGAGCGCGCGCGCGCGCGCGCGCG 1849
 Db 10281 AAGCGGCACTCTGAGACCGGAGCGAGTACCCGACCGCGCGCGCTTCCGCGCGAGTGG 10340
 Oy 1850 CGCTGCGCGCTGTGCGGTGTGTGGGTGTAGACCGGAGTGTCCCGCGCATAGCA 1909
 Db 10341 CGCTGCAACCGGCTGTGAGGCACTGGGGGATGCGCGCGCGCGCGCGCGCGCGCG 10400
 Oy 1910 TCGGTGAGCTGTGTGCTGCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1969
 Db 10401 TGGGGAATGTGGCG 10460
 Oy 1970 TGGTGGCTGCG 2029
 Db 10461 TGGTGGCG 10520
 Oy 2030 TCGAGCGCGCGGAGCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2089
 Db 10521 TCGCGGCGCGGAGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10580
 Oy 2090 TCGCGCGGCTGCAACGCTCCGAGACGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2149
 Db 10581 TCGCGCGGCTGCAACG 10640
 Oy 2150 CGATGCGGCGGAGTGGCG 2209
 Db 10641 AGGCGCGCGGAGTGGCG 10700
 Oy 2210 CGTTCACTCAACGCTCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2269
 Db 10701 CTTTCACTCAACCGCTCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10760

Oy 2270 TGAAGTACCGGCGCGCGCTGCAATGCTCTGGTCAAGCAATGAGCGGAG---GCTTGA 2326
 Db 10761 TGGCTTACCGGGAACCGGAGCTGAGCGTCTGTCTCAAGTGTACGCGGCGCGCGCGCG 10820
 Oy 2327 CAGACGAGTGAAGTTCGCGCGGCTATTGGGTGCGCGCGCGCGCGGAGAGTGTGCTTGG 2386
 Db 10821 CCGGTGAATCAACG 10880
 Oy 2387 CGGATGAGTGAAGCGCGCTGCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2446
 Db 10881 CGGACGCGGTTCG 10940
 Oy 2447 AATGAGCGTGTGCG 2506
 Db 10941 ACGGCGTGTGTGCG 11000
 Oy 2507 CATGTGCG 2566
 Db 11001 ---CGGCGATCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11057
 Oy 2567 GGGCGGTGTGCG 2626
 Db 11058 GGGCG 11115
 Oy 2627 CGTGGCCGACGTAACCTTGGGAGCGCGGAGCGCTACTGAGTGAACAGAAAGCGGAGCG 2686
 Db 11116 GGAAGCG 11175
 Oy 2687 CGGCGGTGAGCG 2746
 Db 11176 CCGCTACTGTGCG 11235
 Oy 2747 CGGTGCG 2806
 Db 11236 CTGGGACCGGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11295
 Oy 2807 GCCGAGAAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2848
 Db 11296 CCACCGGAGCG 11337

Search completed: October 3, 2003, 12:27:59
 Job time : 1144.88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 11:40:41 ; Search time 15260.8 Seconds
(without alignments)
11435.853 Million cell updates/sec

Title: US-10-014-717-1_COPY_7610_11875

Perfect score: 4266

Sequence: 1 gtcggcgagatcgctccatcga.....gggaatcatcgtccatata 4266

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_bc:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_stb:*
12: gb_by:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_stb:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rdi:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_gy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4266	100.0	68750	1 AF210843	AF210843 Sorangium
2	4266	100.0	68750	6 AR193029	AR193029 Sequence
3	4266	100.0	68750	6 AR199551	AR199551 Sequence
4	4266	100.0	68750	6 AR199559	AR199559 Sequence
5	4266	100.0	68750	6 AR199567	AR199567 Sequence
6	4266	100.0	68750	6 AR201097	AR201097 Sequence
7	4266	100.0	68750	6 AR208671	AR208671 Sequence
8	4187.6	98.2	58733	1 AF211789	AF211789 Sorangium
9	4187.6	98.2	71989	6 AR172664	AR172664 Sequence
10	1287.4	30.2	16124	6 AX024277	AX024277 Sequence
11	928.8	21.8	10910	6 AX024276	AX024276 Sequence
12	859	20.1	48000	1 POL505006	POL505006 Sorangium
13	844.4	19.8	49736	1 AF319998	AF319998 Stigmatel
14	817.2	19.2	66808	1 SAU421825	SAU421825 Stigmatel
15	814.2	19.1	3978	6 AR266876	AR266876 Sequence
16	814.2	19.1	12249	6 AR266931	AR266931 Sequence
17	814.2	19.1	18331	6 AR266948	AR266948 Sequence
18	812	19.0	42603	1 AF188287	AF188287 Stigmatel
19	805	18.9	27541	6 AX211706	AX211706 Sequence
20	805	18.9	123580	1 AF263912	AF263912 Streptomy
21	805	18.9	125401	6 AX211739	AX211739 Sequence
22	803.2	18.8	33529	6 AR166425	AR166425 Sequence
23	790.6	18.5	290850	1 SCO939127	SCO939127 Streptomy
24	784	18.4	84985	1 SNA278573	SNA278573 Streptomy
25	772	18.1	113193	1 AF357202	AF357202 Streptomy
26	772	18.1	113193	6 AX703543	AX703543 Sequence
27	770.8	18.1	5355	6 AX697895	AX697895 Sequence
28	770.8	18.1	60196	6 AX697977	AX697977 Sequence
29	766.6	18.0	20394	1 SNA132222	SNA132222 Streptomy
30	766.6	18.0	20394	6 AX067996	AX067996 Sequence
31	766.4	18.0	39314	1 SGR300302	SGR300302 Streptomy
32	765.8	18.0	69644	1 AY179507	AY179507 Streptomy
33	761.2	17.8	104326	1 AB070940	AB070940 Streptomy
34	761.2	17.8	281450	1 AP005032	AP005032 Streptomy
35	755.2	17.7	43280	1 SFU78289	SFU78289 Streptomy
36	745.8	17.5	53784	1 AMU223012	AMU223012 Amycolato
37	745.8	17.5	53789	6 A69720	A69720 Sequence
38	745.8	17.5	90445	1 AF040570	AF040570 Amycolato
39	741.4	17.4	318050	1 BX248344	BX248344 Mycobacte
40	740.8	17.4	16096	1 AB007124	AB007124 Mycobacte
41	740.8	17.4	24292	1 MTCY24G1	MTCY24G1 Mycobacte
42	732.8	17.2	78210	1 AB070949	AB070949 Streptomy
43	732.8	17.2	300425	1 AP005022	AP005022 Streptomy
44	726.2	17.0	65140	6 AX211705	AX211705 Sequence
45	725	17.0	49377	6 I88042	I88042 Sequence

ALIGNMENTS

RESULT 1
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epochlone biosynthesis gene
cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
SOURCE Polyangium cellulosum
ORGANISM Bacteria: Proteobacteria: Deltaproteobacteria: Myxococcales;
Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milamow,M.,
Nowak-Thompson,B., Engel,N., Toupet,C., Strattmann,A., Cyr,D.D.,

TITLE
The biosynthetic gene cluster for the microtubule-stabilizing
agents epothilones A and B from *Sorangium cellulosum* So ce90
JOURNAL
Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE
20130945
PUBMED
10626695
REFERENCE
2 (bases 1 to 68750)
AUTHORS
Molnar, I.
TITLE
Direct Submision
Submitted (03-DEC-1999) Natural Product Genetics, Novartis
Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
12257, Research Triangle Park, NC 27709, USA
FEATURES
source
location/Qualifiers
1..68750
/organism="Sorangium cellulosum"
/mol_type="genomic DNA"
/strain="So ce90"
/db_xref="taxon:56"
<1..1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASIDALFAPATSAVLDGNGHATBRHYLAARAGIEDRALREH
LRIORGSPFHCICGLDTEVELADHDOPLASISFPHARSLRHPDWTSDMLVDGPAV
RLUARGAPLREYEEERERARTAOEARMLAAPCFAPDLPRFEDNAGLELP
MSPEVAERRLRASAYATPELCAALMLGTGAGWSGPYEMLPENLLIGFGLPT
ALAAAPGTSSEALRGARLPASWEVSSKSGOUNIPEALMERLRTIYAMGNADN
LSRFERAEALIAEVRRLRAOPAPPAAGLAIVAGVSGRLGLVTDGALVSGGND
TYMROGRISPVILACTDPPEELAPRLSOMLVANAKGTSKYLTEGSPILYANR
QARPSLVARCFMAYNONAMPDPERGAPFYVQSTIMEFHPTRCLHEPAGSFS
LACDEHLWCESLNGRLMWRPHRPAERFAYLGHPAATWYPSLTNATHVL
MADPRRALIGYDKRTGVEPIVLAETRHPPAVSEDRDIFALTGQPSRDWHEHIR
SASTVADYQOLMDRPMVLNRGLFTTNDRLITLARS"
complement(1900..3171)
CDS
/note="ORF2; similar to hypotheical proteins from
Mycobacterium tuberculosis and Streptomyces coelicolor and
to Dd-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAAFGCALGAEEOGPGDAGALAPAREVMAEVA
AQMPGAWLVARGDGDVADVAVTELGSGAPMRBDTIFRIASMTKAVTAVMMLVE
EGKLDLDSFVDMLPELANKYLAIDGPIDETPAERPIYVDLMTFTMGISFDA
SSPIORALDELGLVNAQPVMTPHGDDEWIRLGTLPKHOGAOMMTVSGISFVA
VGRADOGFPAFVREIRILALGMRDTPFVDPDKLARFAGCGYPTDDEOTGEKTRMD
GASAVASBPAPSPGAGIVSTVDYLLFARMLMNGVHEGRLLSASVREKTYDHL
TAAQKASFFPGFFETHGWMGMAVTPADAVSEVPKYGMDGGGTISWINDPGRGL
IGIVTQSAGLFSGALREFMRSVYVATSA"
3415..5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGLTEROVLLSTVLTALILVTARASGELARLROPEVIGELFG
GVLPSPVVGALAPGFHRLPOEPVGVLSGIVSMGALLMLLIGIEDVAILKEKA
RGALSALGAIAPPLAAGAFSALVUDRLPBGFLGTYLSTANVIAKULIEBESM
RSTVAVQVTLAAGVSEVAAMVAVMTSSSYGSPALAVASALMLSGFLFWLNGRR
LTHLAMVADATRVSKGVSLVLTFTLAALVTRGLHPLIGFALGVLNLSAPT
NRPLDGVQTLVAGLFAVPFVYLAGRVVNSOLRTPAAGTALLATLTAAKVPA
LGARLGLARGSEALVAVAGLKKGGTDLVAVLVEGLISBEATMVAVVLVTVTA
SPALLTWLEKRAPPOESARLEREERAAVYIPGVERILVVAHALGSPATDYES
TVASRKLCETDITELSTEQQAPGSRPAGASRLARLGRVGTIRQREIRGSG
IGAILPASRDHDLVIGARSPARAGMSFGLQDAIVQAESNVLVVGDPEPARAS
ARRILVPIIGLESPALADLAHVALAMPADEVLVLSAQTDPGAVVMRDEPSRVA
ARSVDEAVFRGRILGVRSRVHVGANPSDETTELARAPVLDLVLCYDNGPLGRL
YIGSTVESVVRSRVVALVVAHCGTREQVR"

CDS
complement(5612..5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="WDKPIGRTRCAIAEGYIPGSGNGPEPMTSHERTCLINASDRDA
QVATIVFSDSDPAPKRYVTPARTTRVNRNDLTPEPIPRDIDYASVIESDAPIV
QHTRLDSROAENLSTIAYTDRE"
6226..6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHVDGRRFRGRIRGHTGLHSLMALACGSGSEKTVQGTRLAP
GADARVTDVDPDAATRLIADVYVHLSPRELAGSERFVWQRPSPSPRRVGVLD
YNADSRGKLAETTPVYANFELLITAEKQSSPQSPSSAAVIGTISVC"
7610..11875
/gene="epoa"
/gene="11875"
/gene="epoa"
/gene="epoa"
/note="PPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADPRRAEDPIAVGASCRLPGLGVIDLSGEFWTLLEGSDTV
GRVAPERWDAAMVDPPDARGKTPVRASPLSDVACPDASFFGISPREALRMPARH
LILEYCHALNMAIAPBALVGTGTGVTGIGSEPEYEAALQATASAEIDHGLGTM
PSVAGRISTYALGRCPCVADTVSSLVVHIALCDSRGSESTALAGGSLVLSLP
STLWLSSTRALADGRCAPASADGREGGAVVAKLSGARADAGLIVLVIRG
SAIHNDGSSGLTVPNGSSOEIVLKRALADGCAASGVYEAAGTTLGDPLEIOA
LNAVYGRDVAVTPLLIGSVKTNLGHPEVAGTIGLKVLSVLSHGOIPALHMOALN
PRISMGDLRTVTPARPPMPDMPTPRAGVSPFMSGTMNAVULEEAPACTPAPPE
RPAELVTSATTSALNDQAARLIDHLETPSOLGDAVEPLATTRSMERLAAAT
SREGRLAALDAAQOQTSPGAVERIADSSROKLAFLFTGGAQTLGKGRGLYDWSAF
REAPDLCVRLPNOEIDRLREVMMAEPASVLAALDQAFLOPALFTEYVLAALWRS
MGVEPELVAGHSIGELVAVAGVFSLEDAVFLVAARGRLQALPAGAVSIEAPEA
DVAAVAPHAASVSLAANAPDOVIVAGQOPVHAIAANMARGARTALVSHVAFS
PLMAPLEAPGRVAVESVYRRPSIVVINSGKCTODEVSSRGVWRVRRVRRVPADG
VKALHAKAGTFVEYGPSTLGLVYKAPMPARALIASAGDEPRTVLAELGLW
AVGLVSWAGLPFSGGRVPLPTFMQERKWIITKADARGRARRPAGADEYEBG
GAVGDSRSARLDHPPESGRREKVEAAGRPRLTIDEGVADHLVLRTERAPG
LGEVEIADAGLSENDVQALAGVPPDLPGKPPPLLGEGCGARIVAVGEGVGLV
VGOPIVIALSAGAFTHVTTSALVLPRLQALSAIEAAMPVAVITAWVLDRIARLOP
GERVLIHAATGCVILAAVOMAOHYGAEVHATNGPEKAYLESIGVRYVDSRSPRFV
NDVSTVQKGMLEPRKRVALLBELGLIAAGVFTTPTATPIARVADAFSMAO
AOHLGKVLTLGDPEVQIRIPTHAGAGSTDRDLDDLAAAPARAAALAEALRTQ
VSOVTRPEIRVGAELFTRLGMSLMAVEENRNIEMSLKIKLSTTFLSTSPALLLA
ONILDALATLSTLSEVAENIRAGVQNDVFSGADQDWEIITAL"
7640..8920
/gene="epoa"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
9236..10201
/gene="epoa"
/note="Region: acetate acyltransferase"
10529..11428
/gene="epoa"
/note="Region: enoyl reductase"
11549..11764
/gene="epoa"
/note="Region: acyl carrier protein"
11872..16104
/gene="epos"
11872..16104
/gene="epos"
/note="BPOS P; contains condensation/heterocyclization,"

cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:672425"
/translation="WTINQLNLELHOGIKLAADGERLQIAKKNALNPILARISEH
KSTILTMRORLPESIVPAPAPERAPPLDIOESYMLGRTGATVSGIHAYRED
CTDLDPRLSARFRAKVAHMDLRAHTLPMVMQVLEPKVDADIEIDLRGSTRREA
RLVSRDMSHRIYDTERPLYHVAVALDEKQRLVYSLINDVLSLIIPEWDL
SEYDEPSTLPLTESTRDYVALLESRKSRKSHQSMYMKRIALEPPTPLPKAD
PSTLKEIFRHEQLPSDSWGLKRRKRGKELPTGVILAFVSIGRMKPEFTL
NITLENRLPVRVNDITGDFTSVTLIDITTRDSFPOKARKIOELMEAMDCDS
GIVQREARVYIGORGALFPVLTSLANQVGTSLQRLGTPLYTSTQTPOLDLH
OLYEHDDVLAMDIVDGVPPDLDDMLAYVFLRLREDEKQVRSGLPAPOLE
ARASANTNALLSEHTLHGLPAARVEOLPMOLAVASAKTITYPELSRSRLRLARL
EOGARPTLVAVMEKMEYVAVLAVESGAVPIDADLPAERHILHDEGEKLV
LTPWLDKLSMPGRIQLLVSEAGVEDGQPPMPITQPSDLAVIYITSGTGLPK
QVMDHRGAVNTIIDLINERFEIGPDRVIALSLSFDSYVDVFGIILAGGTVVPA
SKLRDPAHMAELIEREKTVNNSPALRMVVEHREGSPDSLARSLSLSGMIPV
GLPGLAIRGVSVISLGATASISIGVPRVNDLSMISIPYGRILRNOTPHVD
EALRRPVWPGOLYIGVGLATGYWREDEKTRKSFVHPSTGERLYKTCGLRLPD
GNIERMGEDNOIKLRGRVLEIGEIEETLKHPNVRDVIPVGDANLKLAVVP
EGTRRRAEQDASLTERIDAAHAAEDLSDEGVFKLARHRLDLDEKVPDL
TGODPREAGLVYARRRSVTFLEAPIFVEHRELSLSGSEVPEGLATLPERFVSAG
STPYQTAAYVSGRIEGVEGFIYHFEHRLKLSHGIERGHAVQNDVDEBA
PNLPLVGRIDAIESLYGSSREFFLEGGVMAOLMEAPATNGCGVGFQNFQVR
PVLDRHSDVYVHGMIGRUPROVCTTLOQDSPPRATRGAPRGHQRPAIMLRD
FLRTLPBYMPTVFEIDLPLTSNGVDRKALKERKDTSPRHSHATPRDLIEL
LVAARREVILGELVGLQOSFVDLGATSIHIVMRSLLOKRLREIATITELPOYPLGS
LASGRDSDRLDQRPNMQDVEVARKRRRS"
16251..21749
/gene="epob"
16251..21749
/gene="epos B"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"

Query Match 100.0%; Score 4266; DB 1; Length 68750;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGGGGATTCCTCCATGAGGCGGACCCGAAATCCGATTGCGATGTCGAGACGAGT 60
Db 7610 GTGGGGATTCCTCCATGAGGCGGACCCGAAATCCGATTGCGATGTCGAGACGAGT 7669
QY 61 TGCCGCTGCGCGGTGGCGGTGATCGATCGAGCGGTTCTGAGCGCTCTCGAGGGCTCG 120
Db 7610 TGCCGCTGCGCGGTGGCGGTGATCGATCGAGCGGTTCTGAGCGCTCTCGAGGGCTCG 7729
QY 121 CGCGACACCGTGGCGGAGTCCCGCGACGAGCTGGATGACGACGCTGTGTTGATCC 180
Db 7730 CGCGACACCGTGGCGGAGTCCCGCGACGAGCTGGATGACGACGCTGTGTTGATCC 7789
QY 181 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTCTTGAAGGACGTAGCC 240
Db 7790 GACCCCGATGCCCCGGGGAAGACGCCCGTTACGCGCGCATCTTCTTGAAGGACGTAGCC 7849
QY 241 TGCTTGACGCGCTCTTCTTGCGGATCTCGACTCGGAAAGCGCTGGATGAGACCTTCA 300
Db 7850 TGCTTGACGCGCTCTTCTTGCGGATCTCGACTCGGAAAGCGCTGGATGAGACCTTCA 7909
QY 301 CATCGACTCTTGCTGAGAGTGTGCTGGAGGCGCTGAGAAAGCGCGCATGCTTCATCG 360
Db 7910 CATCGACTCTTGCTGAGAGTGTGCTGGAGGCGCTGAGAAAGCGCGCATGCTTCATCG 7969
QY 361 GCGCTCGTCGTTACGGAAGCGGAGTGTTCATCGGGAATCGGCCCTCGGAATATAGAGCC 420
Db 7970 GCGCTCGTCGTTACGGAAGCGGAGTGTTCATCGGGAATCGGCCCTCGGAATATAGAGCC 8029

QY 421 GCGCTGCGCAAGCAGCGCGCTCCGACAGATCGACTCATGCGGCTGGGACGATG 480
Db 8030 GCGCTGCGCAAGCAGCGCGCTCCGACAGATCGACTCATGCGGCTGGGACGATG 8089
QY 481 CCCACGCTCGGAGCGGCGCCGAATCTGTATCCCTCGGACCTGCGAGGCGCTGTGTCGG 540
Db 8090 CCCACGCTCGGAGCGGCGCCGAATCTGTATCCCTCGGACCTGCGAGGCGCTGTGTCGG 8149
QY 541 GTGATACGGCTTATTCGCTCTCGCTGGTGGCGCTTACATGAGCTGTGAGCTTGGC 600
Db 8150 GTGATACGGCTTATTCGCTCTCGCTGGTGGCGCTTACATGAGCTGTGAGCTTGGC 8209
QY 601 TCCGGGATGCTCTCACGCGCCCTGCTGTGGTGGGATTCGATGTTGTGTCGCGACGACC 660
Db 8210 TCCGGGATGCTCTCACGCGCCCTGCTGTGGTGGGATTCGATGTTGTGTCGCGACGACC 8269
QY 661 CTGCTGTGCTCTCGAAGACCCGCGCGCTGGCCAGGACGCTCGCTGCAAGCATTTTCG 720
Db 8270 CTGCTGTGCTCTCGAAGACCCGCGCGCTGGCCAGGACGCTCGCTGCAAGCATTTTCG 8329
QY 721 GCGAGGCGCATGAGGTTTCGACGAGGCGAAGGAGCGCGCTGCTGCTCAAGCGGCTC 780
Db 8330 GCGAGGCGCATGAGGTTTCGACGAGGCGAAGGAGCGCGCTGCTGCTCAAGCGGCTC 8389
QY 781 AGTGAAGCCCGCGCGGACGCGGATCGGATATTTGGCGGTGATTCAGAGATCCCGCATCA 840
Db 8390 AGTGAAGCCCGCGCGGACGCGGATCGGATATTTGGCGGTGATTCAGAGATCCCGCATCA 8449
QY 841 CACGACGCTGCGACAGCGGCTCTGACCGTGGCGAAGCGGACCTCCAGAAATGTGCTG 900
Db 8450 CACGACGCTGCGACAGCGGCTCTGACCGTGGCGAAGCGGACCTCCAGAAATGTGCTG 8509
QY 901 AAACGGCGCTTGGGCGACGCGGCTGGCGCGCTCTTGGGAGTTATGTGAGGCAAC 960
Db 8510 AAACGGCGCTTGGGCGACGCGGCTGGCGCGCTCTTGGGAGTTATGTGAGGCAAC 8569
QY 961 GGCACGGGACGACGCTTGTGACCCCATCGAAATCCAGCTCGAATGCGGTATACGCG 1020
Db 8570 GGCACGGGACGACGCTTGTGACCCCATCGAATCCAGCTCGAATGCGGTATACGCG 8629
QY 1021 CTCGGGCGAGTGTGCCACGCGCGCTGTGATCGGAGTGGTGAAGACCACTTGGCCAT 1080
Db 8630 CTCGGGCGAGTGTGCCACGCGCGCTGTGATCGGAGTGGTGAAGACCACTTGGCCAT 8689
QY 1081 CCTGAGTATGCTGGGGATTCATGCGGCTGCTGAAGGCTGCTTGTCCCTTACACGCG 1140
Db 8690 CCTGAGTATGCTGGGGATTCATGCGGCTGCTGAAGGCTGCTTGTCCCTTACACGCG 8749
QY 1141 CAGATTCTTGCGACCTTCCAGCGCGGACGCTGAACCCCGGATCTCATGGGATGATCTT 1200
Db 8750 CAGATTCTTGCGACCTTCCAGCGCGGACGCTGAACCCCGGATCTCATGGGATGATCTT 8809
QY 1201 CGGCTGACCGTCAAGCGCGCCCGGACACCGTGGCGGACCTGGAATACGCGCGACGCGG 1260
Db 8810 CGGCTGACCGTCAAGCGCGCCCGGACACCGTGGCGGACCTGGAATACGCGCGACGCGG 8869
QY 1261 GGGGTGAGCTGTTCGGCATGAGCGGGAACCAAGCGCACTGTGTCTGGAAGAGCGCGG 1320
Db 8870 GGGGTGAGCTGTTCGGCATGAGCGGGAACCAAGCGCACTGTGTCTGGAAGAGCGCGG 8929
QY 1321 GCGCGACGTCACACGCGCGCGCGCGGACGCGGACGCGGACGCTGTGTCGTCGCA 1380
Db 8930 GCGCGACGTCACACGCGCGCGCGCGGACGCGGACGCGGACGCTGTGTCGTCGCA 8989
QY 1381 AGGACCGGTCAGCCCTTGATGACACAGGCGCGCGCTGCGGACATTTGAGACCTTAC 1440
Db 8990 AGGACCGGTCAGCCCTTGATGACACAGGCGCGCGCTGCGGACATTTGAGACCTTAC 9049
QY 1441 CTTTGGAGTGTCTGGGCGATGTGGCGTTGATGCTGCGACGACGCGCGCATGAGAG 1500
Db 9050 CTTTGGAGTGTCTGGGCGATGTGGCGTTGATGCTGCGACGACGCGCGCATGAGAG 9109

Db	11270	CCGAGCGGAGGTCGCTGTCCTTTGGAGAGACTCTCGCGCTGATTCCGGGAGGCGGTTC	11322
Qy	3721	ACCCCTCCCCCATCGCGAGCGCTCCGATGCCCCTGTGTCCCGATGCGTTCGCGAGCATG	3780
Db	11330	ACCCCTCCCCCATCGCGAGCGCTCCGATGCCCCTGTGTCCCGATGCGTTCGCGAGCATG	11388
Qy	3781	GCGAGGCGGAGCATCTTTGGGAAAGCTGTACTACGCTGGGTGATACCCGAGGCTCGAGTC	3840
Db	11390	GCGAGGCGGAGCATCTTTGGGAAAGCTGTACTACGCTGGGTGATACCCGAGGCTCGAGTC	11448
Qy	3841	CGATTTCGCAACCCACGAGCGCGCGGCCCGCTTCACCGGGGATGGGACCTGTCTCGACAG	3900
Db	11450	CGATTTCGCAACCCACGAGCGCGCGGCCCGCTTCACCGGGGATGGGACCTGTCTCGACAG	11508
Qy	3901	CTCGCGTCACTGTCGCCCGCGCGCGCGCGCGCGCTGAGGCGTTCTCTCGTACGACG	3960
Db	11510	CTCGCGTCACTGTCGCCCGCGCGCGCGCGCGCGCTGAGGCGTTCTCTCGTACGACG	11568
Qy	3961	GTCCTCGAGGTCGTGGCGACACGCCGGAATACAGGTGCGGCGGAGAGCGCTTCAACCCG	4020
Db	11570	GTCCTCGAGGTCGTGGCGACACGCCGGAATACAGGTGCGGCGGAGAGCGCTTCAACCCG	11628
Qy	4021	CTCGGCAATGCACTCGCTCATGAGCCGTGGAGCTGCGCAATGTATTCGAGCGGACCTCAAG	4080
Db	11630	CTCGGCAATGCACTCGCTCATGAGCCGTGGAGCTGCGCAATGTATTCGAGCGGACCTCAAG	11688
Qy	4081	CTGAAGCTGTTCGACGACGCTTCTGTGCCAGCTCCCCCAATATGCGCTTTGTGGCCAAAC	4140
Db	11690	CTGAAGCTGTTCGACGACGCTTCTGTGCCAGCTCCCCCAATATGCGCTTTGTGGCCAAAC	11748
Qy	4141	CTGTGGATGTCTCTCGGCAACAGCTCTCTCTTTGAGGCGGTGGCGCGGAGAACCTACGG	4200
Db	11750	CTGTGGATGTCTCTCGGCAACAGCTCTCTCTTTGAGGCGGTGGCGCGGAGAACCTACGG	11808
Qy	4201	GCAGGCGGCAAAACGACTTCTGTTCATTCGAGCGCAGATCAAGACTGGAAATCATTTGCC	4260
Db	11810	GCAGGCGGCAAAACGACTTCTGTTCATTCGAGCGCAGATCAAGACTGGAAATCATTTGCC	11868
Qy	4261	CTATGATG 4266	
Db	11870	CTATGATG 11875	
RESULT 2			
LOCUS	ARI93029	68750 bp	DNA linear PAT 20-APR-2002
DEFINITION	Sequence 1 from patent US 6346404.		
ACCESSION	ARI93029		
VERSION	ARI93029.1	GI:20238994	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 68750)		
AUTHORS	Schupp,T., Ligon,J.,Madison, Molnar,I., Zirkle,R., Cyr,D.Dawn. and Gorlach,J.		
TITLE	Genes for the biosynthesis of epoethiones		
JOURNAL	Patent: US 6346404-A 1 12-FEB-2002;		
FEATURES	Location/Qualifiers		
source	1..68750		
	/organism="Unknown"		
BASE COUNT	9596 a 22456 c 25539 g 11159 t		
ORIGIN			
Query Match	100.0%; Score 4266; DB 6; Length 68750;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 4266; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Db	7610	GTGGCGGATCTCCATCTGAGCGCGGACGGAATATCGATTGCGATCTCGAGGAGAT	7669
Qy	61	TGCGGCTGCGCGGAGCGGATGATCTGAGCGGATTCTTGAGACCTCTTCGAGGCGCTG	120

Db	7670	TGCGCTCTGCCGATGGCGGTGATGCACTGAGCGGGTTCTGAGCGCTCTCTGAGGGCTCG	7729
Qy	121	CGCGACACCGTCGGGCGAGTCCCGCGGAA CGCTGGATGCAAGCGGTGTTGATCCC	180
Db	7730	CGCGACACCGTCGGGCGAGTCCCGCGGAA CGCTGGATGCAAGCGGTGTTGATCCC	7789
Qy	181	GACCCCGATGCCCGGGGAAAGACGCCCGGTTACGGCGCGCATCTTTCTGAGCGAGTGAACC	240
Db	7790	GACCCCGATGCCCGGGGAAAGACGCCCGGTTACGGCGCGCATCTTTCTGAGCGAGTGAACC	7849
Qy	241	TGCTTCGACGCGCTCTTCTTCGCGATCTCGCATCTGCGAAGCGCTGCGGATGAGACCTTGCA	300
Db	7850	TGCTTCGACGCGCTCTTCTTCGCGATCTCGCATCTGCGAAGCGCTGCGGATGAGACCTTGCA	7909
Qy	301	CATGACCTCTTGTCTGAGAGTGTGTGAGAGCGCTGAGAAACGCGCGCATGCTTCATCG	360
Db	7910	CATGACCTCTTGTCTGAGAGTGTGTGAGAGCGCTGAGAAACGCGCGCATGCTTCATCG	7969
Qy	361	GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGGCCCGTCGAAATAGAGGCC	420
Db	7970	GCGCTCGTCGGTACGGAAACGGGAGTGTTCATCGGGATCGGGCCCGTCGAAATAGAGGCC	8029
Qy	421	GCGCTCGCGCAAGCGGACGGCGTCGGCAGAGATCGACCGCTATGGCGGGCTGGGACGATG	480
Db	8030	GCGCTCGCGCAAGCGGACGGCGTCGGCAGAGATCGACCGCTATGGCGGGCTGGGACGATG	8089
Qy	481	CCGACGTCGGAGCGGGCCGAAATCTCGTATGCGCTTCGGGCTGCGAGGGCGGTGTGTGCG	540
Db	8090	CCGACGTCGGAGCGGGCCGAAATCTCGTATGCGCTTCGGGCTGCGAGGGCGGTGTGTGCG	8149
Qy	541	GTGCAATACGCTATTTGTCTCTGCTGTGGCGCTTATCTGGCTGTCAAGCTTGTGCG	600
Db	8150	GTGCAATACGCTATTTGTCTCTGCTGTGGCGCTTATCTGGCTGTCAAGCTTGTGCG	8209
Qy	601	TCCGGGGGAATGCTCAAGCGCCCTGCGGTGTGGGGTATCGCGATGTTGTGCGCGAGCAC	660
Db	8210	TCCGGGGGAATGCTCAAGCGCCCTGCGGTGTGGGGTATCGCGATGTTGTGCGCGAGCAC	8269
Qy	661	CTCGTGTGCTCTTCGAAGACCCGCGCGCTGGCCAGGACGGTCGCTCAAGGCAATTTTTCG	720
Db	8270	CTCGTGTGCTCTTCGAAGACCCGCGCGCTGGCCAGGACGGTCGCTCAAGGCAATTTTTCG	8329
Qy	721	GCGGAGGCGCATGGGTTTCGACGAGCGGAAAGGTGTGCGCGTCTGTGTCTTCAAGCGCTC	780
Db	8330	GCGGAGGCGCATGGGTTTCGACGAGCGGAAAGGTGTGCGCGTCTGTGTCTTCAAGCGCTC	8389
Qy	781	AGTGAAGCCCGCGCGGGAACGGAGATCGGAATATGCGCGGTGATTCGAGGATCCGGATCAAT	840
Db	8390	AGTGAAGCCCGCGCGGGAACGGAGATCGGAATATGCGCGGTGATTCGAGGATCCGGATCAAT	8449
Qy	841	CACGACGGTGCAGACAGCGCTTGACACCGTGCAGAACGGGAGCTCCCAAGAAATCGTCTG	900
Db	8450	CACGACGGTGCAGACAGCGCTTGACACCGTGCAGAACGGGAGCTCCCAAGAAATCGTCTG	8509
Qy	901	AAACGGGCGCTTGCGGACGACGAGCTGTGCGCGCTCTTCGTGGGTTATGTTCGAGCACAC	960
Db	8510	AAACGGGCGCTTGCGGACGACGAGCTGTGCGCGCTCTTCGTGGGTTATGTTCGAGCACAC	8569
Qy	961	GCGACGGGCAAGACGCTTGTGTGACCCCATCGAAATCCAAAGTCTGAAATGCGGTATACGGC	1020
Db	8570	GCGACGGGCAAGACGCTTGTGTGACCCCATCGAAATCCAAAGTCTGAAATGCGGTATACGGC	8629
Qy	1021	CTCGGGCGAGATGTTCGACACGCGCTGTCTGTATCGGTTCGGTGAAGACCAACTTGGCGAT	1080
Db	8630	CTCGGGCGAGATGTTCGACACGCGCTGTCTGTATCGGTTCGGTGAAGACCAACTTGGCGAT	8689
Qy	1081	CCTGAGTATGCGTGGGGATCTACTGGGCTGTGAAGGTCTGTGTCTTCACACGGG	1140
Db	8690	CCTGAGTATGCGTGGGGATCTACTGGGCTGTGAAGGTCTGTGTCTTCACACGGG	8749
Qy	1141	CAGATTCTTGCGACCTTCCAGCGCCAGGCGTGAACCCCGGATCTCATGGGGTGATCTT	1200

Db 8750 CAGATTCTGGGCACTCCACGCGGAGCGCTGAACCCCGGATCTCATGGGGTATCTT 8809
Qy 1201 CGGCTGACCTGACCGCGCGCCCGGACACCTGTGCGGACTGGAATACGCGCGACGCGCG 1260
Db 8810 CGGCTGACCTGACCGCGCGCCCGGACACCTGTGCGGACTGGAATACGCGCGACGCGCG 8869
Qy 1261 GGGGGTGTGCTGTTGGGCAATGAGCGGGACCAACGCGCAGTGTGTCTGGAAGAAGCGCG 1320
Db 8870 GGGGGTGTGCTGTTGGGCAATGAGCGGGACCAACGCGCAGTGTGTCTGGAAGAAGCGCG 8929
Qy 1321 GCGGCAAGTGCACACCGCGCGCGCGGACGACCGGACAGCTGTGTGTCTGCGCA 1380
Db 8930 GCGGCAAGTGCACACCGCGCGCGCGGACGACCGGACAGCTGTGTGTCTGCGCA 8989
Qy 1381 AGGACCGGCTGACGCTTGATGACAGCGCGCGCGGCTGTGTGTGTGTGTGTGTGTGT 1440
Db 8990 AGGACCGGCTGACGCTTGATGACAGCGCGCGCGGCTGTGTGTGTGTGTGTGTGTGT 9049
Qy 1441 CTTGGCAGTGTCTGGGCGATGTGGCGTTCACTGTGGCGACGACGCGCGCGATGAG 1500
Db 9050 CTTGGCAGTGTCTGGGCGATGTGGCGTTCACTGTGGCGACGACGCGCGATGAG 9109
Qy 1501 CACCGGCTGTGCGGCTGTGCGGACGCTGAGGAGGAGGCTGTGCGGACGCTGTGCGG 1560
Db 9110 CACCGGCTGTGCGGCTGTGCGGACGCTGAGGAGGAGGCTGTGCGGACGCTGTGCGG 9169
Qy 1561 GCGGAGGAGCAGACGCTGTGCGGCTGTGCGGACGCTGTGCGGACGCTGTGCGGACG 1620
Db 9170 GCGGAGGAGCAGACGCTGTGCGGCTGTGCGGACGCTGTGCGGACGCTGTGCGGACG 9229
Qy 1621 CTGCGCTTCTCTTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
Db 9230 CTGCGCTTCTCTTCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9289
Qy 1681 GATGTATGTGTGCGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
Db 9290 GATGTATGTGTGCGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9349
Qy 1741 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
Db 9350 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9409
Qy 1801 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
Db 9410 CTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9469
Qy 1861 CTGTGCGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1920
Db 9470 CTGTGCGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9529
Qy 1921 GTGGCTGTGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1980
Db 9530 GTGGCTGTGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9589
Qy 1981 CGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2040
Db 9590 CGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9649
Qy 2041 GAGGCGGATGT 2100
Db 9650 GAGGCGGATGT 9709
Qy 2101 AACGCTCCGAGACCAAGT 2160
Db 9710 AACGCTCCGAGACCAAGT 9769
Qy 2161 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2220
Db 9770 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9829
Qy 2221 CGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2280
Db 9830 CGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9889

Qy 2281 CGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2340
Db 9890 CGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9949
Qy 2341 TCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400
Db 9950 TCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10009
Qy 2401 GCGCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2460
Db 10010 GCGCTGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10069
Qy 2461 GCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2520
Db 10070 GCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10129
Qy 2521 GCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2580
Db 10130 GCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10189
Qy 2581 CTGGTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640
Db 10190 CTGGTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10249
Qy 2641 CTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2700
Db 10250 CTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10309
Qy 2701 CGCGGTGTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2760
Db 10310 CGCGGTGTCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10369
Qy 2761 GACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2820
Db 10370 GACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10429
Qy 2821 GAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2880
Db 10430 GAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10489
Qy 2881 CTGCGTCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2940
Db 10490 CTGCGTCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10549
Qy 2941 GACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3000
Db 10550 GACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10609
Qy 3001 CTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3060
Db 10610 CTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10669
Qy 3061 GCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3120
Db 10670 GCGGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10729
Qy 3121 GAGGCGT 3180
Db 10730 GAGGCGT 10789
Qy 3181 CTGCTGCGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3240
Db 10790 CTGCTGCGGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10849
Qy 3241 GACAGATATGCGCGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3300
Db 10850 GACAGATATGCGCGCTTCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10909
Qy 3301 GTGGGTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3360
Db 10910 GTGGGTCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10969

OY	3361	GGCAGGCCCGAGAAAGCGGCTCACTCTGGAATCGCTGGGCGGTGGGATATGAGCGCAATTC	3420
Db	10970	GGCAGGCCCGAGAAAGCGGCTCACTGGAATCGCTGGGCGGTGGGATATGAGCGCAATTC	11029
OY	3421	CGCTCGGACCGGATTTCGTCCGCGAAGTCCGCGCGGTGAGACGGCCGACGAGGAGTAGACGTC	3480
Db	11030	CGCTCGGACCGGATTTCGTCCGCGAAGTCCGCGCGGTGAGACGGCCGACGAGGAGTAGACGTC	11089
OY	3481	GTCGTCAACTTCGCTCTTGGGCGAGCTGATGAGCAAGATTTCATCTCTGGGATGCGAC	3540
Db	11090	GTCGTCAACTTCGCTCTTGGGCGAGCTGATGAGCAAGATTTCATCTCTGGGATGCGAC	11149
OY	3541	GGCGCGATTGTTGAGAGCTCGGCAAGCGCGACCTGTTAAGCGGATTAACAGGCTCGGCGTCGGCG	3600
Db	11150	GGCGCGATTGTTGAGAGCTCGGCAAGCGCGACCTGTTAAGCGGATTAACAGGCTCGGCGTCGGCG	11209
OY	3601	CCGTTTCCTGGGCAATCTCTCTCTTCGCTGGTGAATCTCGGCGGAGTGAATGCTCGAGCGCG	3660
Db	11210	CCGTTTCCTGGGCAATCTCTCTCTTCGCTGGTGAATCTCGGCGGAGTGAATGCTCGAGCGCG	11269
OY	3661	CCGGCGCGGGTCCGTGCGCTCTTGAAGAGCTCTTGCGCTGATCCGCGCAGCGCTGTTC	3720
Db	11270	CCGGCGCGGGTCCGTGCGCTCTTGAAGAGCTCTTGCGCTGATCCGCGCAGCGCTGTTC	11329
OY	3721	ACCCCTCCCCCAATCGGAGACGCTCCGGAATGCGCCGTGTCCGGAATGCGTTCGCGAGCAAG	3780
Db	11330	ACCCCTCCCCCAATCGGAGACGCTCCGGAATGCGCCGTGTCCGGAATGCGTTCGCGAGCAAG	11389
OY	3781	GCGCAGGCGGAGCACTCTTGGGAAAGCTGTACTACGCTGGGATGACCCGAGAGTCCAGATC	3840
Db	11390	GCGCAGGCGGAGCACTCTTGGGAAAGCTGTACTACGCTGGGATGACCCGAGAGTCCAGATC	11449
OY	3841	CGTATTCGAACCCACGAGGCGCGCGCGCTGTCAACCGGAGATCGGAGCTTGCTCGACAG	3900
Db	11450	CGTATTCGAACCCACGAGGCGCGCGCGCGCTGTCAACCGGAGATCGGAGCTTGCTCGACAG	11509
OY	3901	CTCGCGTCACTGTCGGCGCGCGCGCGCGCGCGCGCTGTGAGGCGTTCTCCGTACGCGAG	3960
Db	11510	CTCGCGTCACTGTCGGCGCGCGCGCGCGCGCGCGCGCTGTGAGGCGTTCTCCGTACGCGAG	11569
OY	3961	GTCGCGAGGTCGTGCGCAGCGCCGCAAAATCAAGTCCGCGCGGAGCGCTGTTCACCCGC	4020
Db	11570	GTCGCGAGGTCGTGCGCAGCGCCGCAAAATCAAGTCCGCGCGGAGCGCTGTTCACCCGC	11629
OY	4021	CTCGGCAATGGAATCTGCTCATAGCCCGTGAAGCTCGCAATCGTATCGAGCGAGACTCAAG	4080
Db	11630	CTCGGCAATGGAATCTGCTCATAGCCCGTGAAGCTCGCAATCGTATCGAGCGAGACTCAAG	11689
OY	4081	CTGAAGCTGTTCGAGAGAGTTCCTGTCCAGCTCCCCCAATATCGGCTTTGTTGGCCCAAAAC	4140
Db	11690	CTGAAGCTGTTCGAGAGAGTTCCTGTCCAGCTCCCCCAATATCGGCTTTGTTGGCCCAAAAC	11749
OY	4141	CTGTTGATGTCCTCGGCAAGGCTCTCTCTTGAAGCGGGTGGCGGCGAGAACCTAAGCG	4200
Db	11750	CTGTTGATGTCCTCGGCAAGGCTCTCTCTTGAAGCGGGTGGCGGCGAGAACCTAAGCG	11809
OY	4201	GCAGGCGGTGCAAAAGCACTTGTCTCACTCGGCGCAGATCAAGACTGGGAAATCAATTGCC	4260
Db	11810	GCAGGCGGTGCAAAAGCACTTGTCTCACTCGGCGCAGATCAAGACTGGGAAATCAATTGCC	11869
OY	4261	CTATGGA 4266	
Db	11870	CTATGGA 11875	

RESULT 3	AR199551	68750 bp	DNA	linear	PAT 20-APR-2002
LOCUS	AR199551				
DEFINITION	Sequence 1 from patent US 6355457.				
ACCESSION	AR199551				
VERSION	AR199551.1	GI:20249625			
KEYWORDS					
SOURCE	Unknown.				

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	SOURCE
Unknown.	1 (bases 1 to 68750)	Schupp,T., Ligon,J.,Madison, Molnar,I., Zirkle,R., Cyr,D.,Dawn, and Gorlach,J.	Genes for the biosynthesis of epochnones	Patent: US 6355457-A 1 12-MAR-2002;	Location/Qualifiers	1..68750
						/organism="unknown"
BASE COUNT	9596 a	22456 c	25539 g	11159 t		
ORIGIN						

Query Match	100.0%;	Score 4266;	DB 6;	Length 68750;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 4266;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	I	GTGGCGGATCGTCCCATATGAGCGCCGACCCGAAAGTTCGGAATTGGCATTCGTCGGAGCGAAT	60
Db	7610	GTGGCGGATCGTCCCATATGAGCGCCGACCCGAAAGTTCGGAATTGGCATTCGTCGGAGCGAAT	766
QY	61	TGCCGTCGCGCGGATGCGGTGATTCATCTGAGCGGATTCGTGAGCGTCTCTCGAGGGCTCG	120
Db	7670	TGCCGTCGCGCGGATGCGGTGATTCATCTGAGCGGATTCGTGAGCGTCTCTCGAGGGCTCG	772
QY	121	CGCGACACCGTCGAGCGAGTCCCGCGGAAACGCTGGAGTGCAGCAGCGTGTGATCCC	180
Db	7730	CGCGACACCGTCGAGCGAGTCCCGCGGAAACGCTGGAGTGCAGCAGCGTGTGATCCC	778
QY	181	GACCCCGATGCCCCCGGGGGAAGAAGCCCGCTTACGGCGCATCTTTCTGAGCGACGTAAGCC	240
Db	7790	GACCCCGATGCCCCCGGGGGAAGAAGCCCGCTTACGGCGCATCTTTCTGAGCGACGTAAGCC	784
QY	241	TGCTTCGACGCGCTCCTCTTCCTCGGCATCTTCGCTCGGAGCGCTGCGAGTGAACCTCTGA	300
Db	7850	TGCTTCGACGCGCTCCTCTTCCTCGGCATCTTCGCTCGGAGCGCTGCGAGTGAACCTCTGA	790
QY	301	CATCGACTTCCTGCTGAGAGTGTGCTGCGAGCGCTGAGAAACGCCCGCATTCGTCATATG	360
Db	7910	CATCGACTTCCTGCTGAGAGTGTGCTGCGAGCGCTGAGAAACGCCCGCATTCGTCATATG	796
QY	361	GCGCTCGTCGCGTAACGGAACGGGAGTGTTCATCTGGGATCGGCCCGCTCGCAATATGAGGCC	420
Db	7970	GCGCTCGTCGCGTAACGGAACGGGAGTGTTCATCTGGGATCGGCCCGCTCGCAATATGAGGCC	802
QY	421	GCGCTCGCGCAAGCGAGCGCGCTCGCGAAGATCGACTCATATGCGCGGCTGGGGACGATG	480
Db	8030	GCGCTCGCGCAAGCGAGCGCGCTCGCGAAGATCGACTCATATGCGCGGCTGGGGACGATG	808
QY	481	CCGACGTCGGAAGCGGGCGGAAATCTTCGTAATGCCCTCGGCTGCGAGGGCGGTGTGTCGG	540
Db	8090	CCGACGTCGGAAGCGGGCGGAAATCTTCGTAATGCCCTCGGCTGCGAGGGCGGTGTGTCGG	814
QY	541	GTGGAATACGGCTATTCGTCCTCGCTGAGCGCTTTCATCTGGCTGTGACAGACTTGC	600
Db	8150	GTGGAATACGGCTATTCGTCCTCGCTGAGCGCTTTCATCTGGCTGTGACAGACTTGC	820
QY	601	TCCGGGGAATGCTCCACGCGCCCTGCGTGGTGGGGTATATCGCTGATGTTGTGCGCCGACACC	660
Db	8210	TCCGGGGAATGCTCCACGCGCCCTGCGTGGTGGGGTATATCGCTGATGTTGTGCGCCGACACC	826
QY	661	CTCGGTATGCTCTCGAAGAACCCCGGGCGCTGGCGACGAGAACGTCGCTGCAAGGCAATTTGG	720
Db	8270	CTCGGTATGCTCTCGAAGAACCCCGGGCGCTGGCGACGAGAACGTCGCTGCAAGGCAATTTGG	832
QY	721	GCGAGGCGCGATGGGTTTCGAGACGAGCGGAAGGTCGCGTGTGTCTCTCAACGCGCTC	780
Db	8330	GCGAGGCGCGATGGGTTTCGAGACGAGCGGAAGGTCGCGTGTGTCTCTCAACGCGCTC	838
QY	781	AGTGAAGCCCGCGCGAGACGCGCATTCGGAATTTGGGCGGTATTCGAGATCCCGCGATCAAT	840
Db	8390	AGTGAAGCCCGCGCGAGACGCGCATTCGGAATTTGGGCGGTATTCGAGATCCCGCGATCAAT	844

OY	841	CACGACGGGTGGAGGACGAGGGGCTGTGACCGTGGCGGAA	CGGGAGCTCCCAAGAAATCGTGTG	900		
Db	8450	CACGACGGGTGGAGGACGAGGGGCTGTGACCGTGGCGGAA	CGGGAGCTCCCAAGAAATCGTGTG	85099		
OY	901	AAACGGGGCCCTGGCGGGA	CGCAGCGCTGCGCGCTTTCGGTGGGTATATGT	CGAGGACAC	960	
Db	8510	AAACGGGGCCCTGGCGGGA	CGCAGCGCTGCGCGCGCTTTCGGTGGGTATATGT	CGAGGACAC	85659	
OY	961	GGCACGGGCGACGAGCGCTTGGTGA	CCCATCGAATCCAA	GCTCTGAATGCGGTATACGGC	1020	
Db	8570	GGCACGGGCGACGAGCGCTTGGTGA	CCCATCGAATAATCCAA	GCTCTGAATGCGGTATACGGC	8629	
OY	1021	CTCGGGCGAGATATATCGGACGCGCGCTGTGTATCGGGTGTG	GAAGAACCAACTTGGCCAT	1080		
Db	8630	CTCGGGCGAGATATATCGGACGCGCGCGCTGTGTATCGGGTGTG	GAAGAACCAACTTGGCCAT	8689		
OY	1081	CCTGAGTATGAGGTGCGGGAGATCACTGGGCTGTGCTGAAGAGT	CGTCTTGTCCCTTCAGCACGGG	1140		
Db	8690	CCTGAGTATGAGGTGCGGGAGATCACTGGGCTGTGCTGAAGAGT	CGTCTTGTCCCTTCAGCACGGG	8749		
OY	1141	CAGATTCTCTGGCGACCTTCACGCGCGAGGCGCTGAA	CCCCCGATCTCA	ATGAGGATCTT	1200	
Db	8750	CAGATTCTCTGGCGACCTTCACGCGCGAGGCGCTGAA	CCCCCGATCTCA	ATGAGGATCTT	8809	
OY	1201	CGGCTGACCGGTCA	CGCGCGGCGCCCGGACACCGGTGGCCGGAC	CTGGAATATGCCCGGACGCGCGC	1260	
Db	8810	CGGCTGACCGGTCA	CGCGCGGCGCCCGGACACCGGTGGCCGGAC	CTGGAATATGCCCGGACGCGCGC	8869	
OY	1261	GGGGTGAAGCTGTTGCGGATGAG	CGGGACCAACCGCGCACGTGTGATCTGGAAGAGGCGCG	1320		
Db	8870	GGGGTGAAGCTGTTGCGGATGAG	CGGGACCAACCGCGCACGTGTGATCTGGAAGAGGCGCG	8929		
OY	1321	GCGGCGACGTGCA	CA	CCGCGCGGCGCGGAGCGGACCGGCGA	GACTGTGTCTGTCCGCA	1380
Db	8930	GCGGCGACGTGCA	CA	CCGCGCGGCGCGGAGCGGACCGGCGA	GACTGTGTCTGTCCGCA	8989
OY	1381	AGGACCGGCGGTCA	CGCCCTGGAATGCA	ACGCGCGCGCGCGCTGCGCGACACATCTTGA	AGACTTAC	1440
Db	8990	AGGACCGGCGGTCA	CGCCCTGGAATGCA	ACGCGCGCGCGCGCTGCGCGACACATCTTGA	AGACTTAC	9049
OY	1441	CCTTCGACGATGTCTTGGGCGATATGTGGCGTTCA	GTGCGGAGGACGCGGAGCGCGATGGAG	1500		
Db	9050	CCTTCGACGATGTCTTGGGCGATATGTGGCGTTCA	GTGCGGAGGACGCGGAGCGCGATGGAG	9109		
OY	1501	CACCGGCTCGCGGTGGCGGACGTCGA	GGGAGGGGCTGCGGCGACCCCTTGA	ACGCTGCG	1560	
Db	9110	CACCGGCTCGCGGTGGCGGACGTCGA	GGGAGGGGCTGCGGCGACCCCTTGA	ACGCTGCG	9169	
OY	1561	GGCGACGGGACAGACGTGCGCCGGGTGCGGTGCG	CACTATCGCCGATTCTCTCACGCGGCAAG	1620		
Db	9170	GGCGACGGGACAGACGTGCGCCGGGTGCGGTGCG	CACTATCGCCGATTCTCTCACGCGGCAAG	9229		
OY	1621	CTCGGCTTTCCTTCA	CCGAGCGGACGAGGAGGCGGACGCTGGGGCAT	GGGCGGTGAC	1680	
Db	9230	CTCGGCTTTCCTTTCCTTCA	CCGAGCGGAGGAGGAGGCGGACGCTGGGGCAT	GGGCGGTGAC	9289	
OY	1681	GATGTATGTGTCGCGGTTCCGCGAGGCGCTTTCGA	CCCTGTGCGTGAAGGCTGTTCA	CCAGGAG	1740	
Db	9290	GATGTATGTGTCGCGGTTCCGCGAGGCGCTTTCGA	CCCTGTGCGTGAAGGCTGTTCA	CCAGGAG	9349	
OY	1741	CTTCGACCGGCGCTCTCGCGAGGTGATGTGGGCGGAA	CCGGCGCAGGCTGACGCGCGCGCTG	1800		
Db	9350	CTTCGACCGGCGCTCTCGCGAGGTGATGTGGGCGGAA	CCGGCGCAGGCTGACGCGCGCGCGCTG	9409		
OY	1801	CTTCGACCGAGACGCTTAC	CCCGAGCGCGGCGCTGTTCA	CTTTCGAATATGCGCTGCGCGC	1860	
Db	9410	CTTCGACCGAGACGCTTAC	CCCGAGCGCGGCGCTGTTCA	CTTTCGAATATGCGCTGCGCGC	9469	
OY	1861	CTGTGCGCGGTGCTGTGGGATGTAGAGCGCGGAGTTGAT	TCGCGCGGACATAGCATCGTGA	CGCTG	1920	
Db	9470	CTGTGCGCGGTGCTGTGGGATGTAGAGCGCGGAGTTGAT	TCGCGCGGACATAGCATCGTGA	CGCTG	9529	
OY	1921	GTTGGCTGCTGCTGCGGCGGCGTGTTCCTGCTTGA	GACGCGGCTTTCCTGTGCGCTGCG	1980		

Db	9530	GTGGCTTCCTGCGCTGGCGGGCGGTTCCTCGCTTGAAGACGCGGGTGTTCCTGGAGGCTGGCG	9589
Qy	1981	CGCGGGCGCCTGATGCAAGCGCTGCGGCGCGCGCGGGCGATGGTGTGCATTCGAAGCGCGCG	2040
Db	9590	CGCGGGCGCCTGATGCAAGCGCGCTGCGGCGCGCGCGGGCGATGGTGTGCATTCGAAGCGCGCGCG	9649
Qy	2041	GAGGCGCGAATGGCTCTCGCGGGTGGGGCGCGCAACGAGACGTCGGGTGTGCATTCGCGCGGGTTC	2100
Db	9650	GAGGCGCGAATGGCTCTCGCGGGTGGGGCGCGCGCAAGCGTCGGGTGTGCATTCGCGCGGGTTC	9709
Qy	2101	AACGCTCCGGAACCAAGGTGTGCATTCGCGGGCGCGCGGCAACCCGTGCATCGATTCGCGCGCG	2160
Db	9710	AACGCTCCGGAACCAAGGTGTGCATTCGCGGGCGCGCGGCAACCCGTGCATTCGATTCGCGCGCG	9769
Qy	2161	GCGATGGCGCGCGCGGGGGCGGACCAAGCGCTCCACGTCTCGCATGCTTCCACTCA	2220
Db	9770	GCGATGGCGCGCGCGGGGGCGGACCAAGCGCTCCACGTCTCGCATGCTTCCACTCA	9829
Qy	2221	CCGCTCATGGCCCCCGATTCCTGAGAGGGTTCGGGCGGTGTGGCCGAGTCGTGGAGCTACCGG	2280
Db	9830	CCGCTCATGGCCCCCGATTCCTGAGAGGGTTCGGGCGGTGTGGCCGAGTCGTGGAGCTACCGG	9889
Qy	2281	CGGCGCGTCGATTCGTCTCGGTGACGAATCTGAGGCGGAGAGGCTTGCACAGACGAGTGAGC	2340
Db	9890	CGGCGCGTCGATTCGTCTCGGTGACGAATCTGAGGCGGAGAGGCTTGCACAGACGAGTGAGC	9949
Qy	2341	TGCGCGGGCTAATTTGGGTGCGCACAAGCGGAGAGTGTGTCGCTTCGCGGATGGAGTGAAAG	2400
Db	9950	TGCGCGGGCTAATTTGGGTGCGCACAAGCGGAGAGTGTGTCGCTTCGCGGATGGAGTGAAAG	10005
Qy	2401	GCGCTGCAAGCGGCGCGGTGTGGGGCACCTTCGTGTGAGGTGCGTTCGGAATTCGACGCTGCTC	2460
Db	10010	GCGCTGCAAGCGGCGCGGTGTGGGGCACCTTCGTGTGAGGTGCGTTCGGAATTCGACGCTGCTC	10065
Qy	2461	GGCGCTGGTGTCTGCTGTGCATAGCGGAGCGCCCGGCGCGGCTGCTGCATTCGTCGCGCGCT	2520
Db	10070	GGCGCTGGTGTCTGCTGTGCATAGCGGAGCGCCCGGCGCGGCTGCTGCATTCGTCGCGCGCT	10122
Qy	2521	GGGCGTGCACAGCGCGGCGACCGGTCTCGAGGCGCTCGGCGGCTCTGGGCGCGTCCGTCCGTGCG	2580
Db	10130	GGGCGTGCACAGCGCGGCGACCGGTCTCGAGGCGCTCGGCGGCGCTCTGGGCGCGTCCGTCCGTGCG	10188
Qy	2581	CTGGTCTCTCTGGGCGCGGCTCTTCCCTCAAGGGGGGCGCGGTCGCCGTGCCACAGTAC	2640
Db	10190	CTGGTCTCTCTGGGCGCGGCTCTTCCCTCAAGGGGGGCGCGGTCGCCGTGCCACAGTAC	10245
Qy	2641	CTTTGGAGGCGAGAGGCGCTACCTGAGTTCGACAGGAAGCCGACGACGCGGCGCGTGGCGAC	2700
Db	10250	CTTTGGAGGCGAGAGGCGCTACCTGAGTTCGACAGGAAGCCGACGACGCGGCGCGTGGCGAC	10305
Qy	2701	CGCGGTGCTCCGGGAGCGGCTTCACGACGAGGTTCGAGAGGGGGGCGCGGTGCGCGCGCG	2760
Db	10310	CGCGGTGCTCCGGGAGCGGCTTCACGACGAGGTTCGAGAGGGGGGCGCGGTGCGCGCGCG	10365
Qy	2761	GACCGGCGACGCGCTCGGCTTCGACATTCGCGCGCCCGGAGCGGAGCGGAGGAGGTC	2820
Db	10370	GACCGGCGACGCGCTCGGCTTCGACATTCGCGCGCCCGGAGCGGAGCGGAGGAGGTC	10422
Qy	2821	GAGGCGCGCGGAGCGGTCGTTCCGGTTCGAGATTCGATGAGGCGAGCGGTCCTGATTCAC	2880
Db	10430	GAGGCGCGCGGAGCGGTCGTTCCGGTTCGAGATTCGATGAGGCGAGCGGTCCTGATTCAC	10485
Qy	2881	CTCGTGTCTTGGGTCACGAGCGGCGCGCCCTTGATTCGGGCGAGGTCGAGATTCGCGCTC	2940
Db	10490	CTCGTGTCTTGGGTCACGAGCGGCGCGCCCTTGATTCGGGCGAGGTCGAGATTCGCGCTC	10545
Qy	2941	GACGCGCGCGGAGCTCAGCTTCAAATGATTCAGCTGCGCTGTGGCATGTGTCCCGACGAC	3000
Db	10550	GACGCGCGCGGAGCTCAGCTTCAAATGATTCAGCTGCGCTGTGGCATGTGTCCCGACGAC	10605
Qy	3001	CTGCGCGGGAAGCCCAACCTTCGCGTGTGCTGTGAGAGCGAGTGTGCGCGGGCGCATTCGTC	3060

Db 8090 CCCAGCTCGAGCGGGCCGAATCTCGTATGCCCCCTCGGGCTCGAGAGGCCGTGTGCGG 8149
Qy 541 GTGATACGGGCTTATTTGTCCTCGCTGATGAGCCGTTTCATCTGACCCTGTGCAGAGCTTGGCG 600
Db 8150 GTGATACGGGCTTATTTGTCCTCGCTGATGAGCCGTTTCATCTGACCCTGTGCAGAGCTTGGCG 8209
Qy 601 TCCCGGGAATGCTCCACCGGCGCTGCTGATGAGGATTCGCTGATGAGGATTCGCTGATGAGG 660
Db 8210 TCCCGGGAATGCTCCACCGGCGCTGCTGATGAGGATTCGCTGATGAGGATTCGCTGATGAGG 8269
Qy 661 CTGCTGATGCTCGAAGAGCCGCGGCTGAGCAGAGGAGCTGCTGCTGAGGAGCATTTTTCG 720
Db 8270 CTGCTGATGCTCGAAGAGCCGCGGCTGAGCAGAGGAGCTGCTGCTGAGGAGCATTTTTCG 8329
Qy 721 GCGGAGGCGGATGAGGTTGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
Db 8330 GCGGAGGCGGATGAGGTTGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8389
Qy 781 AGTGAAGCCCGGCGGAGCGGAGATCGGATATTTGCGGATATTCAGAGATCCCGGATCAT 840
Db 8390 AGTGAAGCCCGGCGGAGCGGAGATCGGATATTTGCGGATATTCAGAGATCCCGGATCAT 8449
Qy 841 CACGACGCTGCGAGCAGCGGCTGAGCCTGAGCCGAAAGGAGGCTCCCAAGAAATCGTCTG 900
Db 8450 CACGACGCTGCGAGCAGCGGCTGAGCCTGAGCCGAAAGGAGGCTCCCAAGAAATCGTCTG 8509
Qy 901 AAACGGGCGCTGCGGAGCGCAGGAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 8510 AAACGGGCGCTGCGGAGCGCAGGAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8569
Qy 961 GGCAGCGGAGCGAGCCTTGTGATACCCCATCGAAATTCAGAGCTCTGATATCGGATATACGCG 1020
Db 8570 GGCAGCGGAGCGAGCCTTGTGATACCCCATCGAAATTCAGAGCTCTGATATCGGATATACGCG 8629
Qy 1021 CTCGGGCGAGATGTCGACCGCGCTGATCGGCTGAGTGAAGCAACCTTGGCAT 1080
Db 8630 CTCGGGCGAGATGTCGACCGCGCTGATCGGCTGAGTGAAGCAACCTTGGCAT 8689
Qy 1081 CCTGAGTATGCTGCGGAGATCACTGAGGCTGCTGAGAGTGTCTTGTCTCTTGCAGACGGG 1140
Db 8690 CCTGAGTATGCTGCGGAGATCACTGAGGCTGCTGAGAGTGTCTTGTCTCTTGCAGACGGG 8749
Qy 1141 CAGATTTCTGCGGACCTTCCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
Db 8750 CAGATTTCTGCGGACCTTCCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8809
Qy 1201 CGGCTGAGCCGTCAGCGGCGCCGAGCACCGTGGCCGAGCTGAGATATACGCGGAGCGGCG 1260
Db 8810 CGGCTGAGCCGTCAGCGGCGCCGAGCACCGTGGCCGAGCTGAGATATACGCGGAGCGGCG 8869
Qy 1261 GGGGATGAGCTGCTTGGGCACTGAGCGGAGCAACGCGGACGCTGCTGAGAGAGGCGCG 1320
Db 8870 GGGGATGAGCTGCTTGGGCACTGAGCGGAGCAACGCGGACGCTGCTGAGAGAGGCGCG 8929
Qy 1321 GGGGAGGAGTGAACCGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
Db 8930 GGGGAGGAGTGAACCGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 8989
Qy 1381 AGGACCGGCTCAGCCCTGAGTACAGAGCGGCGCGGCTGCGGACCATCTGAGAGACTTAC 1440
Db 8990 AGGACCGGCTCAGCCCTGAGTACAGAGCGGCGCGGCTGCGGACCATCTGAGAGACTTAC 9049
Qy 1441 CCTTGGCAGTGTCTGGGCGATGATGAGCTTCACTTGGCGACAGCGGCGAGCCGATGAG 1500
Db 9050 CCTTGGCAGTGTCTGGGCGATGATGAGCTTCACTTGGCGACAGCGGCGAGCCGATGAG 9109
Qy 1501 CACCGGCTCGCGGTGCGGCGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Db 9110 CACCGGCTCGCGGTGCGGCGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9169
Qy 1561 GCGCAGGAGCAGAGCTGCGCGGCTGCGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Db 9170 GCGCAGGAGCAGAGCTGCGCGGCTGCGGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9229

Qy 1621 CTGCGCTTTCTCTTACCGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
Db 9230 CTGCGCTTTCTCTTACCGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 9289
Qy 1681 GATG 1740
Db 9290 GATG 9349
Qy 1741 CTGACCGGCGCTTCGCGAGAGTGTAGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCG 1800
Db 9350 CTGACCGGCGCTTCGCGAGAGTGTAGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCGAGCCG 9409
Qy 1801 CTGACCGAGACGCTTACCCAGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db 9410 CTGACCGAGACGCTTACCCAGCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9469
Qy 1861 CTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
Db 9470 CTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9529
Qy 1921 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Db 9530 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9589
Qy 1981 GCGGCGCGCTGATGACAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
Db 9590 GCGGCGCGCTGATGACAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9649
Qy 2041 GAGGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
Db 9650 GAGGCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9709
Qy 2101 AACGCTCCGAGCAGAGTGTGATGCGGAGCGCGGAGCAACCGTGTGATGATGATGATGATG 2160
Db 9710 AACGCTCCGAGCAGAGTGTGATGCGGAGCGCGGAGCAACCGTGTGATGATGATGATGATG 9769
Qy 2161 GCGATGCG 2220
Db 9770 GCGATGCG 9829
Qy 2221 CCGCTCATGCG 2280
Db 9830 CCGCTCATGCG 9889
Qy 2281 CCGCGCTGATGCTCTGCTGATGAGCTTGAAGCGGAAAGGCTTGTGACAGAGAGGATGAGC 2340
Db 9890 CCGCGCTGATGCTCTGCTGATGAGCTTGAAGCGGAAAGGCTTGTGACAGAGAGGATGAGC 9949
Qy 2341 TCGCGCGGCTATTTGGGTCGCGCAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400
Db 9950 TCGCGCGGCTATTTGGGTCGCGCAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 10009
Qy 2401 GCGCTGACAGCGGCGCGGTCGCGGCACTTGTGAGGAGTGTGCTGCGAAATTCAGCGCTGCTC 2460
Db 10010 GCGCTGACAGCGGCGCGGTCGCGGCACTTGTGAGGAGTGTGCTGCGAAATTCAGCGCTGCTC 10069
Qy 2461 GCGCTGAGCTGCTGCTGATGCGGAGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 2520
Db 10070 GCGCTGAGCTGCTGCTGATGCGGAGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTG 10129
Qy 2521 GGGGCTGACAGAGCGGCGAGCGCTGCTGAGAGCGCTGCGGCGGCGCTGCGGCGGCTGCGG 2580
Db 10130 GGGGCTGACAGAGCGGCGAGCGCTGCTGAGAGCGCTGCGGCGGCGCTGCGGCGGCTGCGG 10189
Qy 2581 CTGCTCTCTGAGCGCGGCTCTTCCCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2640
Db 10190 CTGCTCTCTGAGCGCGGCTCTTCCCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10249
Qy 2641 CTTTGGAGCGGAGCGCTACTGATGACAGAGAAAGCGGAGAGCGGCGCGCTGCTGCGAGC 2700
Db 10250 CTTTGGAGCGGAGCGCTACTGATGACAGAGAAAGCGGAGAGCGGCGCGCTGCTGCGAGC 10309

2701 GCGCGTCTCCGGAGCGGCTGACGACGAGTGCAGAGAGGGGGGCGCGGTGCGCGCGG 2760
Db GCGCGTCTCCGGAGCGGCTGACGACGAGTGCAGAGAGGGGGGCGCGGTGCGCGCGG 10369
2761 GACCGGCGACGCTCCGGCTGACCAATCCGCGCGCGCGGAGCGCGCGCGGAGAAAGT 2820
Db GACCGGCGACGCTCCGGCTGACCAATCCGCGCGCGCGGAGCGCGCGGAGAAAGT 10429
2821 GAGCGCGCGCGACCGCTCCGGCTGACCAATCCGCGCGCGGAGCGCGCGGAGAAAGT 2880
Db GAGCGCGCGCGACCGCTCCGGCTGACCAATCCGCGCGCGGAGCGCGCGGAGAAAGT 10489
2881 CTCTGCTTCCGGCTGACCAATCCGCGCGCGGAGCGCGCGGAGAAAGT 2940
Db CTCTGCTTCCGGCTGACCAATCCGCGCGCGGAGCGCGCGGAGAAAGT 10549
2941 GACCGGCGCGGCTCAGCTTCAATGATCCAGCTCGCGCGGAGCGCGCGGAGAAAGT 3000
Db GACCGGCGCGGCTCAGCTTCAATGATCCAGCTCGCGCGGAGCGCGCGGAGAAAGT 10609
3001 CTGCGCGGAGAAAGCGCAACCTCCGCTGCTCGAGGCGAGTGGCGCGCGCGCTGCT 3060
Db CTGCGCGGAGAAAGCGCAACCTCCGCTGCTCGAGGCGAGTGGCGCGCGCGCTGCT 10669
3061 GCGCGGCGAGGCGGTGAACCGCTCGTGGTGGGCAACCGGTGATCGCTTCCGCG 3120
Db GCGCGGCGAGGCGGTGAACCGCTCGTGGTGGGCAACCGGTGATCGCTTCCGCG 10729
3121 GAGCGCTTGTCTACCAAGTCAACGCTGCGCTGGCTGGCTGCTGCTGCTGCTGCTG 3180
Db GAGCGCTTGTCTACCAAGTCAACGCTGCGCTGGCTGGCTGCTGCTGCTGCTGCTG 10789
3181 CTCTCGCGGATCGAGGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 3240
Db CTCTCGCGGATCGAGGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 10849
3241 GACGAGATAGCTCCGCTTCAACCGCGGAGCGCGGTGCTGATTCAGCGCGAGCGCG 3300
Db GACGAGATAGCTCCGCTTCAACCGCGGAGCGCGGTGCTGATTCAGCGCGAGCGCG 10909
3301 GTCCGTCTCCGCGCGGTGCAAGTGGCGCGAGCACTGGAGCGCGAGTTCAGCGCG 3360
Db GTCCGTCTCCGCGCGGTGCAAGTGGCGCGAGCACTGGAGCGCGAGTTCAGCGCG 10969
3361 GGCAGCGCGGAGAAAGCGCGCTTCAACCGCGGAGCGCGGTGCTGATTCAGCGCG 3420
Db GGCAGCGCGGAGAAAGCGCGCTTCAACCGCGGAGCGCGGTGCTGATTCAGCGCG 11029
3421 CGCTCGGACCGGCTTCTGCGCGAGCTGCGCGGTGGAAGCGCGCGGAGTGAAGT 3480
Db CGCTCGGACCGGCTTCTGCGCGAGCTGCGCGGTGGAAGCGCGCGGAGTGAAGT 11089
3481 GTGCTCAACTGCTTCCGCGGAGCTGATTCAGCAAGAGTTCATCTCTCGAGTCGAC 3540
Db GTGCTCAACTGCTTCCGCGGAGCTGATTCAGCAAGAGTTCATCTCTCGAGTCGAC 11149
3541 GGCAGCGGCTTGGAGCTGCGGAGCGAGCTGTAACGCGGTAACCACTGCGGCTGCG 3600
Db GGCAGCGGCTTGGAGCTGCGGAGCGAGCTGTAACGCGGTAACCACTGCGGCTGCG 11209
3601 CGCTTCTGCGCAATCT 3660
Db CGCTTCTGCGCAATCT 11269
3661 CCGCGCGCGGCTTCT 3720
Db CCGCGCGCGGCTTCT 11270
3721 ACCCTTCCCGCATCGAGCGCTCCGATCGCGGCTGCGCGGATCGGTTCCGAGCAT 3780
Db ACCCTTCCCGCATCGAGCGCTCCGATCGCGGCTGCGCGGATCGGTTCCGAGCAT 11330
3781 GCGCAGCGCGAGCATCTTGGAGAGCTCGTCACTCACTGAGTGAACCGAGGCTCCAGAT 3840

11390 GCGCAGCGCGAGCATCTTGGAGAGCTGTAAGTCTGAGTGAACCGAGGCTCAGATC 11449
Db GCGCAGCGCGAGCATCTTGGAGAGCTGTAAGTCTGAGTGAACCGAGGCTCAGATC 11449
3841 GCGATTCACACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3900
Db GCGATTCACACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11509
3901 CTGCGCTGAGCTGCG 3960
Db CTGCGCTGAGCTGCG 11569
3961 GTCTCGAGGCTGCG 4020
Db GTCTCGAGGCTGCG 11629
4021 CTGCGGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4080
Db CTGCGGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11689
4081 CTGAGCTGCTGAGCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4140
Db CTGAGCTGCTGAGCGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11749
4141 CTGTTGATGCTTCTGCGCAAGCTTCTCTCTGAGGCGGCTGCGCGGAGAACTT 4200
Db CTGTTGATGCTTCTGCGCAAGCTTCTCTCTGAGGCGGCTGCGCGGAGAACTT 11809
4201 GCGAGCGCGGAAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4260
Db GCGAGCGCGGAAAGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11869
4261 CTATGA 4266
Db CTATGA 11875

RESULT 5
ARI99567 68750 bp DNA linear PAT 20-APR-2002
LOCUS ARI99567
DEFINITION Sequence 1 from patent US 6355459.
ACCESSION ARI99567
VERSION ARI99567.1 GI:20249641
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp,T., Iijon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn. and
Gorlach,J.
TITLE Genes for the biosynthesis of epoethilones
JOURNAL Patent: US 6355459-A, 1 12-MAR-2002;
FEATURES
source Location/Qualifiers
1..68750
/organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 4266; DB 6; Length 68750;
Best Local Similarity 100.0%; Ptd. No. 0;
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GTGCGGATGCTTCCATCGAGCGCGAGCGGAGATCGATTGCGATCGTGGAGCGAGT 60
Db GTGCGGATGCTTCCATCGAGCGCGAGCGGAGATCGATTGCGATCGTGGAGCGAGT 7669
61 TGCGGTGCGCGGCGGTGAGTGAAGCGGCTTCTGAGCGCTTCTGAGCGGCTCG 120
Db TGCGGTGCGCGGCGGTGAGTGAAGCGGCTTCTGAGCGCTTCTGAGCGGCTCG 7729
7670 TGCGGTGCGCGGCGGTGAGTGAAGCGGCTTCTGAGCGCTTCTGAGCGGCTCG 7729
121 GCGGACACGCTGCGGCGAGTCCCGCGGAGCGCTGGAGATGAGAGAGTGTGATCC 180
Db GCGGACACGCTGCGGCGAGTCCCGCGGAGCGCTGGAGATGAGAGAGTGTGATCC 7789
7730 GCGGACACGCTGCGGCGAGTCCCGCGGAGCGCTGGAGATGAGAGAGTGTGATCC 7789

Oy	1.8	GACCCGATGCCCCGGGGAAACACGCCGTTACGGCGCATCTTTTCTGAGGACGTAAGCC	240
Db	7790	GACCCCGATGCCCGGGGAAAGACGCCGTTACGGCGCATTTTCTGAGCGACTAGCC	7849
Oy	241	TGCTTTCGACGCCCTCTTTCGGGATCTCGCCTCGAAGCGCTGCGGATGSAACCTTCA	300
Db	7850	TGCTTTCGACGCCCTCTTTCGGGATCTCGCCTCGGAAAGCGCTGCGGATGSAACCTTCA	7909
Oy	301	CATGGACTCTTGTCTGAGAGTGTGCTGGAGGCGCTGGAAACGCCCGATCGCTCAATCG	360
Db	7910	CATGGACTCTTGTCTGAGAGTGTGCTGGAGGCGCTGGAGAAACGCCCGATCGCTCAATCG	7969
Oy	361	GCGCTCCGCGGCTACGGAAACGGGAGTGTTCATCCGGGATCGGCGCGCTCCGAATAGAGGCC	420
Db	7970	GCGCTCCGCGGCTACGGAAACGGGAGTGTTCATCCGGGATCGGCGCGCTCCGAATAGAGGCC	8029
Oy	421	GCGCTCCGCGCAAGCGACCGCGCTCGCAGAGATTCGACGCTCATGCGGGCTTGGGACAGATG	480
Db	8030	GCGCTCCGCGCAAGCGACCGCGCTCCGCAAGATTCGACGCTCATGCGGGCTTGGGACAGATG	8089
Oy	481	CCGACGCTCGGAGCGGGGCCGAATCTCGTATGCCCTTCGGGCTGCGAGGCGCGTGTGTCGG	540
Db	8090	CCGACGCTCGGAGCGGGGCCGAATCTCGTATGCCCTTCGGGCTGCGAGGCGCGTGTGTCGG	8149
Oy	541	GTGGAATAGGCGCTATTCGTCCTCCGTGTGGCGCTTCATCTGCGCTGTCAAGCTTTCGCG	600
Db	8150	GTGGAATAGGCGCTATTCGTCCTCCGTGTGGCGCTTCATCTGCGCTGTCAAGAGTTTCGCG	8209
Oy	601	TCCGGGGGAATCTCCACGGGCGCTGGCTGTGGGGGATTCGCTGATGTTGTGCGCGAGACC	660
Db	8210	TCCGGGGGAATCTCCACGGGCGCTTCGCTGTGGGGGATTCGCTGATGTTGTGCGCGAGACC	8269
Oy	661	CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCACAGGAACGGTGCCTGCAAGGCAATTTTCG	720
Db	8270	CTCGTGTGGCTCTCGAAGACCCGGGCGCTGGCCACAGGAACGGTGCCTGCAAGGCAATTTTCG	8329
Oy	721	GCGAGGCGCGCATTGGGTTTCGACAGAGCGCAAGGGTGCCTGCTGTGTCTTCAACGGCTC	780
Db	8330	GCGAGGCGCGCATTGGGTTTCGACAGAGCGCAAGGGTGCCTGCTGTGTCTTCAACGGCTC	8389
Oy	781	AGTGGAGCCCCCGCGGACCGGCGATTCGATTTTGGCGGGATTTCGAGATTCGCGATCAAT	840
Db	8390	AGTGGAGCCCCCGCGGACCGGCGATTCGATTTTGGCGGGATTTCGAGATTCGCGATCAAT	8449
Oy	841	CACGACGGTGCAGACGCGGCTGTGACCGTGCAGAACGGGAGCTCCCAAGAAATCGTCTG	900
Db	8450	CACGACGGTGCAGACGCGGCTGTGACCGTGCAGAACGGGAGCTCCCAAGAAATCGTCTG	8509
Oy	901	AAACGGGCGCTTCGGCGGACGCAAGCGCTGCGCCCGCTTTCGGTGGGTTATGTGAGGCAAC	960
Db	8510	AAACGGGCGCTTCGGCGGAGCGCAAGCGCTGCGCCCGCTTTCGGTGGGTTATGTGAGGCAAC	8569
Oy	961	GGCACGGGCAAGACGCTTGTGTGACCCCATTCGAAATTCGAAGCTTCGAATTCGGGTATAGGCG	1020
Db	8570	GGCACGGGCAAGACGCTTGTGTGACCCCATTCGAATTCGAAGCTTCGAATTCGGGTATAGGCG	8629
Oy	1021	CTCGGGCAGATGTGTGCACACGCGCTGCTGATCGGTCGTGATGGAAGCAACCACTTGGCCAT	1080
Db	8630	CTCGGGCAGATGTGTGCACACGCGCTGCTGATCGGTCGTGATGGAAGCAACCACTTGGCCAT	8689
Oy	1081	CTTGAGTATGCGTCGGGATCATCTGAGGCTGCTGTAAGATCGTCTTCCCTTACACACGGG	1140
Db	8690	CTTGAGTATGCGTCGGGATCATCTGAGGCTGCTGTAAGATCGTCTTCCCTTACACACGGG	8749
Oy	1141	CAGATTTCTTGCGCACCTTCACGCGCAGGCGGCTGAACCCCGGATTCATGAGGGTGAATCTT	1200
Db	8750	CAGATTTCTTGCGCACCTTCACGCGCAGGCGGCTGAACCCCGGATTCATGAGGGTGAATCTT	8809
Oy	1201	CGGCTGACCGTTCACGCGCGCCCGGACACCGTGTGCGGACTGGAATACGCCCGCGACGGCGC	1260
Db	8810	CGGCTGACCGTTCACGCGCGCCCGGACACCGTGTGCGGACTGGAATACGCCCGCGCGACGGCGC	8869
Oy	1261	GGGGTGAGCTCGTTGCGCATGAGGGGGAACCAACGCGCATGTTGTGCTGGAAGAGCGCGC	1320

Db	8870	GGGGTAGCTCTTTCCGCAATGAGCGGGACCAACGCGAGCTGGTGTCTGGAAAGCGCCG	8929
Qy	1321	GCGGCGACGTGCACACCGCGCGCGGAGCGACCGGACAGACTGCTGGTCTGTGGCA	1380
Db	8930	GGCGCGACGTGCACACCGCGCGCGCGGAGCGACCGGACAGACTGCTGGTGTCTGGCA	8989
Qy	1381	AGGACCGCGGTACAGCCCTGGATGACAGAGCGCGCGCGCTGGCGACCATCTGGAGACCTAC	1440
Db	8990	AGGACCGCGGTACAGCCCTGGATGACAGAGCGCGCGCGCTGGCGACCATCTGGAGACCTAC	9049
Qy	1441	CCCTTCGAGAGTCTCGGGCGGATGTGGCGTTAGTCTGGCGACGACGCGACGCGATGGAG	1500
Db	9050	CCCTTCGAGAGTCTCGGGCGGATGTGGCGTTAGTCTGGCGACGACGCGACGCGATGGAG	9109
Qy	1501	CACCGGCTCGCGGTGGCGGACGCTCGAGGGAGGGGCTGCGGGCAAGCTTGGACGCTGCG	1560
Db	9110	CACCGGCTCGCGGTGGCGGACGCTCGAGGGAGGGGCTGCGGGCAAGCTTGGACGCTGCG	9169
Qy	1561	GGCGAGGGACAGAGAGTGTCCCGGCTGGCTGGTCCGAGATTCGCGCATTCCTACCGCGGACG	1620
Db	9170	GGCGAGGGACAGAGAGTGTCCCGGCTGGCTGGTCCGAGATTCGCGCATTCCTACCGCGGACG	9229
Qy	1621	CTGCGCTTTCTCTTCAACCGGACAGGGGGCGCAGACGCTGGGCGATGGGCGTGGGCTGTAC	1680
Db	9230	CTGCGCTTTCTCTTCAACCGGACAGGGGGCGCAGACGCTGGGCGATGGGCGTGGGCTGTAC	9289
Qy	1681	GATGTATGTGTCCGCGTTCGCGAGGCGGTTGACCTGTGCTGAGGCTGTTCAACAGAGAG	1740
Db	9290	GATGTATGTGTCCGCGTTCGCGAGGCGGTTGACCTGTGCTGAGGCTGTTCAACAGAGAG	9349
Qy	1741	CTGCAACCGCGCGCTCCGCGAGGTGATGTGGCCCGGACCGGCGACGCTTCGACGCGCGCTG	1800
Db	9350	CTGCAACCGCGCGCTCCGCGAGGTGATGTGGCCCGGACCGGCGACGCTTCGACGCGCGCTG	9409
Qy	1801	CTGCAACGAGACGCTTCAACCGGCGCGCGCTGTTCACCTTCGAAATATGGCGCTGCGCG	1860
Db	9410	CTGCAACGAGACGCTTCAACCGGCGCGCGCTGTTCACCTTCGAAATATGGCGCTGCGCG	9469
Qy	1861	CTGTGGCGGTGTGGGGGTGTAGAGCCCGGAGTTGGTCGCGCGGACCATAGCATCGGTAGCTG	1920
Db	9470	CTGTGGCGGTGTGGGGGTGTAGAGCCCGGAGTTGGTCGCGCGGACCATAGCATCGGTAGCTG	9529
Qy	1921	GTGGCTGCTGCTGGTGGCGGCGCTTTCGCTTGAAGACCGCGGTTCCTGTGTGTGCTGCG	1980
Db	9530	GTGGCTGCTGCTGGTGGCGGCGCTTTCGCTTGAAGACCGCGGTTCCTGTGTGTGCTGCG	9589
Qy	1981	CGCGGGGCGCGATGACAGGGGCTGGCGCGCGCGCGCGGCGGATGTGATTCGAGAGCGCG	2040
Db	9590	CGCGGGGCGCGATGACAGGGGCTGGCGCGCGCGCGCGGCGGATGTGATTCGAGAGCGCG	9649
Qy	2041	GAGGCGGATGTGGCTGTGGGTGGCGCGACGACGATCGGTGTGATCGCTCGCGGCTC	2100
Db	9650	GAGGCGGATGTGGCTGTGGGTGGCGCGACGACGATCGGTGTGATCGCTCGCGGCTC	9709
Qy	2101	AAGCCTCCGACACAGGTGTGTATGTGGCGCGCGCGGAGAACCTGTGTGATGTGGCGCG	2160
Db	9710	AAGCCTCCGACACAGGTGTGTATGTGGCGCGCGCGGAGAACCTGTGTGATGTGGCGCG	9769
Qy	2161	GCGATGGCGCGCGCGGGGGCGGAAACGAAGCGCTCAGCGTTCGACATGTGGCTTCATCTCA	2220
Db	9770	GCGATGGCGCGCGCGGGGGCGGAAACGAAGCGCTCAGCGTTCGACATGTGGCTTCATCTCA	9829
Qy	2221	CCGCTCATGTGCCCCGATGTGGAGCGGTTGCGGCGTGTGGCCGATCGGTGAGCTACCGG	2280
Db	9830	CCGCTCATGTGCCCCGATGTGGAGCGGTTGCGGCGTGTGGCCGATCGGTGAGCTACCGG	9889
Qy	2281	CGGCGCTGTGATGCTCTGTGTACGAATCTGAGCGGAGAGGCTTTCACACAGACGAGGTGAC	2340
Db	9890	CGGCGCTGTGATGCTCTGTGTACGAATCTGAGCGGAGAGGCTTTCACACAGACGAGGTGAC	9949
Qy	2341	TGCGCGGGGCTATTTGGGTGGCGCACGGCGGAGAGGTGTGGGCTTCGCGGATGTGAGTAA	2400

Db 9950 TCGCCGGCTATGGGTGGCCACCGCCGAGAGTGTCGCTTCGCGGATGAGTAAAG 10009
Qy 2401 GCGCTGCACGCGCGCGGTGCGGCACTTTCGTGAGTGTGTCGAAATTCAGCGTGTCTC 2460
Db 10010 GCGCTGCACGCGCGCGGTGCGGCACTTTCGTGAGTGTGTCGAAATTCAGCGTGTCTC 10069
Qy 2461 GCGCTGTGCTGCTGCTGATGTCGAGACGCGCGCGGTGCTGTCGATGTCGCGCT 2520
Db 10070 GCGCTGTGCTGCTGCTGATGTCGAGACGCGCGCGGTGCTGTCGATGTCGCGCT 10129
Qy 2521 GCGCGTGCAGAGCGCGCGCACTGTCGAGGCGTTCGCGCGGTGCTGCGCGGTGCTGCG 2580
Db 10130 GCGCGTGCAGAGCGCGCGCACTGTCGAGGCGTTCGCGCGGTGCTGCGCGGTGCTGCG 10189
Qy 2581 CTGCTCTCTGCGCGCGCTTCTTCCCTTCAGAGGCGCGCGGTGCTGCGCGCGCTGAC 2640
Db 10190 CTGCTCTCTGCGCGCGCTTCTTCCCTTCAGAGGCGCGCGGTGCTGCGCGCGCTGAC 10249
Qy 2641 CCTTGGCAGCGCGAGCGCTACTGATTCGACGAAAGCCGACGAGCGCGCGGTGCGAGC 2700
Db 10250 CCTTGGCAGCGCGAGCGCTACTGATTCGACGAAAGCCGACGAGCGCGCGGTGCGAGC 10309
Qy 2701 CGCGCTGCTCCGCGAGCGCGGTTCACGACGAGTTCGAGAGGCGCGCGGTGCGCGCGC 2760
Db 10310 CGCGCTGCTCCGCGAGCGCGGTTCACGACGAGTTCGAGAGGCGCGCGGTGCGCGCGC 10369
Qy 2761 GACCGCGCAGCGCTGCGCTCGACCATCGCGCGCGCGAGAGCGCGCGCGAGAAAGTTC 2820
Db 10370 GACCGCGCAGCGCTGCGCTCGACCATCGCGCGCGCGAGAGCGCGCGCGAGAAAGTTC 10429
Qy 2821 GAGCGCGCGCGAGCGCTGCTGCGGTTCGAGATTCGATGAGCGCGGTTCGATGAC 2880
Db 10430 GAGCGCGCGCGAGCGCTGCTGCGGTTCGAGATTCGATGAGCGCGGTTCGATGAC 10489
Qy 2881 CTGCGTCTTCGCGGTTCACGAGAGCGCGCGCGCGGTTCGAGAGTTCGAGTTCGCGCT 2940
Db 10490 CTGCGTCTTCGCGGTTCACGAGAGCGCGCGCGCGGTTCGAGAGTTCGAGTTCGCGCT 10549
Qy 2941 GACCGCGCGCGGTCTCAGCTTCATGATGTCAGCTGCGCTGCGGTGAGTTCGCGAGC 3000
Db 10550 GACCGCGCGCGGTCTCAGCTTCATGATGTCAGCTGCGCTGCGGTGAGTTCGCGAGC 10609
Qy 3001 CTGCGCGAGAAAGCCCAACCTTCGCTGCTGCTGCGAGCGAGTTCGCGCGCGCTGTC 3060
Db 10610 CTGCGCGAGAAAGCCCAACCTTCGCTGCTGCTGCGAGCGAGTTCGCGCGCGCTGTC 10669
Qy 3061 GCGCGTGCAGAGCGGTGTAAGCGCTGCTGCTGCGCGCAACCGGTATGCTTCGCGCG 3120
Db 10670 GCGCGTGCAGAGCGGTGTAAGCGCTGCTGCTGCGCGCAACCGGTATGCTTCGCGCG 10729
Qy 3121 GAGCGCTTTCGCTACCGAGTTCACGAGTTCGCGTGTGCTGCGAGCGAGTTCGCGCGCT 3180
Db 10730 GAGCGCTTTCGCTACCGAGTTCACGAGTTCGCGTGTGCTGCGAGCGAGTTCGCGCGCT 10789
Qy 3181 CTCTCGCGGATTCGAGCGCGCGCGCAATGCTGCGTTCGAGCGAGTTCGAGCGCTC 3240
Db 10790 CTCTCGCGGATTCGAGCGCGCGCGCAATGCTGCGTTCGAGCGAGTTCGAGCGCTC 10849
Qy 3241 GACAGAAATAGCCCGCTTCAGCGCGCGCGAGCGGTGCTGATTCATGCGCGCGCGCG 3300
Db 10850 GACAGAAATAGCCCGCTTCAGCGCGCGCGAGCGGTGCTGATTCATGCGCGCGCGCGCG 10909
Qy 3301 GTCGCTGCTGCGCGCGGTGAGGCGCGAGCAGCGTGGAGCGCGAGTTCATGCGCGCGCG 3360
Db 10910 GTCGCTGCTGCGCGCGGTGAGGCGCGAGCAGCGTGGAGCGCGAGTTCATGCGCGCGCG 10969
Qy 3361 GGCAGCGCGCGAGAAAGCGCGCTTACCTGAGTTCGCTGCGCGGTGCGGTATTCGAGCTTC 3420
Db 10970 GGCAGCGCGCGAGAAAGCGCGCTTACCTGAGTTCGCTGCGCGGTGCGGTATTCGAGCTTC 11029
Qy 3421 CGCTGCGAGCGGTTCGTGCGCGAGCTGCGCGCGGTGAGCGCGCGCGAGAGTTCAGCTC 3480
Db 11030 CGCTGCGAGCGGTTCGTGCGCGAGCTGCGCGCGGTGAGCGCGCGCGAGAGTTCAGCTC 11089

Qy 3481 GTGCTCACTGCTCTCGCGCGAGTGTATCGACAAAGTTTCAATCTCTCGATTCGAC 3540
Db 11090 GTGCTCACTGCTCTCGCGCGAGTGTATCGACAAAGTTTCAATCTCTCGATTCGAC 11149
Qy 3541 GCGCGTTCGAGAGCTTCGCAAGCGCGAGTTCGCGGATTCACAGCTGCGGTGCGG 3600
Db 11150 GCGCGTTCGAGAGCTTCGCAAGCGCGAGTTCGCGGATTCACAGCTGCGGTGCGG 11209
Qy 3601 CGGTTCGCGAGTTCCTCTCTGCTGCTGAGTTCGCGGAGTTCGCGGAGTTCGAGCGG 3660
Db 11210 CGGTTCGCGAGTTCCTCTCTGCTGCTGAGTTCGCGGAGTTCGCGGAGTTCGAGCGG 11269
Qy 3661 CGCGCGCGGTTCGCTGCTTCGAGAGTTCCTCGCGCTGATTCGCGCGCGGTTC 3720
Db 11270 CGCGCGCGGTTCGCTGCTTCGAGAGTTCCTCGCGCTGATTCGCGCGCGGTTC 11329
Qy 3721 ACCCTTCCTCCATTCGCGAGCTTCGATTCGCGCGGTTCGCGAGTTCGCGAGCTG 3780
Db 11330 ACCCTTCCTCCATTCGCGAGCTTCGATTCGCGCGGTTCGCGAGTTCGCGAGCTG 11389
Qy 3781 GCGCAGCGCGAGTTCCTGAGAGTTCGACGCTGCGGTGAGCGCGAGTTCAGATC 3840
Db 11390 GCGCAGCGCGAGTTCCTGAGAGTTCGACGCTGCGGTGAGCGCGAGTTCAGATC 11449
Qy 3841 GGTATTCCAACCCAGCAGCGCGCGCGCGCTTCACCGCGGATTCGAGCTTCGAGCAG 3900
Db 11450 GGTATTCCAACCCAGCAGCGCGCGCGCGCGCTTCACCGCGGATTCGAGCTTCGAGCAG 11509
Qy 3901 CTGCGTACGCTGCGCGCGCGCGCGCGCGCGCGCGGTTCGAGAGCTTCCTCGATCAG 3960
Db 11510 CTGCGTACGCTGCGCGCGCGCGCGCGCGCGCGCGGTTCGAGAGCTTCCTCGATCAG 11569
Qy 3961 GTCCTGAGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGGTTCGAGAGCTTCCTCGATCAG 4020
Db 11570 GTCCTGAGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGGTTCGAGAGCTTCCTCGATCAG 11629
Qy 4021 CTGCGCATGAGCTGCTCATGAGCGGTGAGCGTTCGATTCGAGCGAGCTTCAG 4080
Db 11630 CTGCGCATGAGCTGCTCATGAGCGGTGAGCGTTCGATTCGAGCGAGCTTCAG 11689
Qy 4081 CTGAAGCTGTCAGAGCTTCCTGTCAGCTTCGCGCGCGCGCGCGGTTCGAGAGCTTCAG 4140
Db 11690 CTGAAGCTGTCAGAGCTTCCTGTCAGCTTCGCGCGCGCGCGCGGTTCGAGAGCTTCAG 11749
Qy 4141 CTGTTGATTCGCTTCGCGAGCTTCCTCTTCGAGCGCGGTTCGAGAGCTTCAG 4200
Db 11750 CTGTTGATTCGCTTCGCGAGCTTCCTCTTCGAGCGCGGTTCGAGAGCTTCAG 11809
Qy 4201 GCGCGCGTGCAGAAAGCACTTCGCTCATCGCGCGCGAGTTCAGAGCTTCAGAGCTTCAG 4260
Db 11810 GCGCGCGTGCAGAAAGCACTTCGCTCATCGCGCGCGAGTTCAGAGCTTCAGAGCTTCAG 11869
Qy 4261 CTATGA 4266
Db 11870 CTATGA 11875

RESULT 6
AR201097 68750 bp DNA linear PAT 20-APR-2002
LOCUS AR201097
DEFINITION Sequence 1 from patent US 6358719.
ACCESSION AR201097
VERSION AR201097.1 GI:20251985
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Schnupp, T., Iigo, J., Madson, J., Molnar, I., Zirkle, R., Cyr, D., Dawn, and
Gorlach, U.
TITLE Genes for the biosynthesis of epoethinones
JOURNAL Patent: US 6358719-A-1 19-MAR-2002;

FEATURES Location/Qualifiers
source 1..68750 /organism="unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN
Query Match 100.0%; Score 4266; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GTGGCGATGTCATCCATCGAGCGCGAGCCGAAGATCCGATTGCGATCGTCGAGCGAGT 60
Db GTGGCGATGTCATCCATCGAGCGCGAGCCGAAGATCCGATTGCGATCGTCGAGCGAGT 7669
61 TCCCGTCTGCGCGGTGCGTGAATCGATCTGACCGGGTCTTGAACGCTCTCGAGGGCTCG 120
Db TCCCGTCTGCGCGGTGCGTGAATCGATCTGACCGGGTCTTGAACGCTCTCGAGGGCTCG 7729
121 CGCGACACCGTGGGGGATCCCGCGCGACGCTGGAGATGACGACGCGTGTGATATCC 180
Db CGCGACACCGTGGGGGATCCCGCGCGACGCTGGAGATGACGACGCGTGTGATATCC 7789
7730 CGCGACACCGTGGGGGATCCCGCGCGACGCTGGAGATGACGACGCGTGTGATATCC 7789
181 GACCCGATGCGCGCGGGAAGAAGCCGTTACGCGCGCATCTTCTCTGAGACGATAGCC 240
Db GACCCGATGCGCGCGGGAAGAAGCCGTTACGCGCGCATCTTCTCTGAGACGATAGCC 7849
241 TGCCTGACGCGCTCTCTTGGGACATCTGCGCTCGGAAGCGCTGCGAATGGAACCTTGA 300
Db TGCCTGACGCGCTCTCTTGGGACATCTGCGCTCGGAAGCGCTGCGAATGGAACCTTGA 7909
301 CATTCACCTTGTGAGAGGTGCTGGAGGCGCTGGAGAGAGCGCGCGATTCGCTCATATG 360
Db CATTCACCTTGTGAGAGGTGCTGGAGGCGCTGGAGAGAGCGCGCGATTCGCTCATATG 7969
7910 CATTCACCTTGTGAGAGGTGCTGGAGGCGCTGGAGAGAGCGCGCGATTCGCTCATATG 7969
361 GCGCTCTGCTGATACGGAAGCGGAGTGTTCATCGAGATGCGCGCTCGGATATAGAGCC 420
Db GCGCTCTGCTGATACGGAAGCGGAGTGTTCATCGAGATGCGCGCTCGGATATAGAGCC 8029
7970 GCGCTCTGCTGATACGGAAGCGGAGTGTTCATCGAGATGCGCGCTCGGATATAGAGCC 8029
421 GCGCTCTGCTGATACGGAAGCGGAGTGTTCATCGAGATGCGCGCTCGGATATAGAGCC 480
Db GCGCTCTGCTGATACGGAAGCGGAGTGTTCATCGAGATGCGCGCTCGGATATAGAGCC 8089
8030 GCGCTCTGCTGATACGGAAGCGGAGTGTTCATCGAGATGCGCGCTCGGATATAGAGCC 8089
481 CCCAGGCTGGAGCGCGCGGAGATCTGTATGCCCTCGGGCTGGAGGCGCGTGTGCGG 540
Db CCCAGGCTGGAGCGCGCGGAGATCTGTATGCCCTCGGGCTGGAGGCGCGTGTGCGG 8149
8090 CCCAGGCTGGAGCGCGCGGAGATCTGTATGCCCTCGGGCTGGAGGCGCGTGTGCGG 8149
541 GTGATACGCGCTATTCGTCTCGCTGCTGAGCGGTTTCATCTGAGCTGTGAGAGCTTGGCC 600
Db GTGATACGCGCTATTCGTCTCGCTGCTGAGCGGTTTCATCTGAGCTGTGAGAGCTTGGCC 8209
8150 GTGATACGCGCTATTCGTCTCGCTGCTGAGCGGTTTCATCTGAGCTGTGAGAGCTTGGCC 8209
601 TCCGCGGAATGCTCCACGCGCGCTGCTGAGGAGTATCGCTGATGTTGTGCGGACACC 660
Db TCCGCGGAATGCTCCACGCGCGCTGCTGAGGAGTATCGCTGATGTTGTGCGGACACC 8269
8210 TCCGCGGAATGCTCCACGCGCGCTGCTGAGGAGTATCGCTGATGTTGTGCGGACACC 8269
661 CTCGTGTGCTCTCGAAGACCGGGGCGTGGCCAGGAGAGTGCCTGCAAGGATTTTGG 720
Db CTCGTGTGCTCTCGAAGACCGGGGCGTGGCCAGGAGAGTGCCTGCAAGGATTTTGG 8329
8270 CTCGTGTGCTCTCGAAGACCGGGGCGTGGCCAGGAGAGTGCCTGCAAGGATTTTGG 8329
721 GCGGAGCGCGATGAGTTGCGAGCGAGCGAAGGAGTGGCGCTGTGCTCTCAAGCGGCTC 780
Db GCGGAGCGCGATGAGTTGCGAGCGAGCGAAGGAGTGGCGCTGTGCTCTCAAGCGGCTC 8389
8330 GCGGAGCGCGATGAGTTGCGAGCGAGCGAAGGAGTGGCGCTGTGCTCTCAAGCGGCTC 8389
781 AGTGAAGCCCGCGCGGAGCGGATCGGATATTTGGCGGTGATTCGAGGATCCGCGATCAAT 840
Db AGTGAAGCCCGCGCGGAGCGGATCGGATATTTGGCGGTGATTCGAGGATCCGCGATCAAT 8449
8390 AGTGAAGCCCGCGCGGAGCGGATCGGATATTTGGCGGTGATTCGAGGATCCGCGATCAAT 8449
841 CACGACGCTGCGAGCGCGCTGACACGCGCGAAGCGGAGCGTCCCAAGAAATTCGCGTGC 900
Db CACGACGCTGCGAGCGCGCTGACACGCGCGAAGCGGAGCGTCCCAAGAAATTCGCGTGC 9509
8450 CACGACGCTGCGAGCGCGCTGACACGCGCGAAGCGGAGCGTCCCAAGAAATTCGCGTGC 9509
901 AACCGGCGCTGCGAGCGCAAGGCTGCGCGCGTCTTCTGAGGAGTTATTCGAGGACAC 960
Db AACCGGCGCTGCGAGCGCAAGGCTGCGCGCGTCTTCTGAGGAGTTATTCGAGGACAC 8569

961 GGCACGGGACAGACGCTTGTGATCCCATCCGAATCCAGCTTGAATGCGGTATACGGC 1020
Db GGCACGGGACAGACGCTTGTGATCCCATCCGAATCCAGCTTGAATGCGGTATACGGC 8629
1021 CTCGGGCGAGATGTCGACCGCGCTGCTGATTCGGGTGCTGTAAGACCAACCTTGGCCAT 1080
Db CTCGGGCGAGATGTCGACCGCGCTGCTGATTCGGGTGCTGTAAGACCAACCTTGGCCAT 8689
8630 CTCGGGCGAGATGTCGACCGCGCTGCTGATTCGGGTGCTGTAAGACCAACCTTGGCCAT 8689
1081 CTTGATATGCTGCGGGGATCACTGGGCTGTGAAGTGTCTTCTCTTCCCTTACGACGGG 1140
Db CTTGATATGCTGCGGGGATCACTGGGCTGTGAAGTGTCTTCTCTTCCCTTACGACGGG 8749
8690 CTTGATATGCTGCGGGGATCACTGGGCTGTGAAGTGTCTTCTCTTCCCTTACGACGGG 8749
1141 CAGATTCCTGCGACCTTCCACGCGGAGCGGCTGAAACCCCGGATTCATATGGGTGATCTT 1200
Db CAGATTCCTGCGACCTTCCACGCGGAGCGGCTGAAACCCCGGATTCATATGGGTGATCTT 8809
8750 CAGATTCCTGCGACCTTCCACGCGGAGCGGCTGAAACCCCGGATTCATATGGGTGATCTT 8809
1201 CGGCTGACCGCTACGCGCGCGCGGACACGCTGAGCGGACTGGAATACCGCGGACGGGCG 1260
Db CGGCTGACCGCTACGCGCGCGCGGACACGCTGAGCGGACTGGAATACCGCGGACGGGCG 8869
8810 CGGCTGACCGCTACGCGCGCGCGGACACGCTGAGCGGACTGGAATACCGCGGACGGGCG 8869
1261 GGGGTGACCTGCTTGGGATGAGCGGAGCAACCGGACGCTGTGCTGAAAGGCGCGCG 1320
Db GGGGTGACCTGCTTGGGATGAGCGGAGCAACCGGACGCTGTGCTGAAAGGCGCGCG 8929
8870 GGGGTGACCTGCTTGGGATGAGCGGAGCAACCGGACGCTGTGCTGAAAGGCGCGCG 8929
1321 GCGGAGACGTGACACCGCGCGCGCGGAGCGGACCGGAGAGCTGCTGTGCGGCA 1380
Db GCGGAGACGTGACACCGCGCGCGCGGAGCGGACCGGAGAGCTGCTGTGCGGCA 8989
8930 GCGGAGACGTGACACCGCGCGCGCGGAGCGGACCGGAGAGCTGCTGTGCGGCA 8989
1381 AGGACCGGCTGACCGCTGATGACAGGCGGCGCGCTGCGGACCATCTGAGAGACTTAC 1440
Db AGGACCGGCTGACCGCTGATGACAGGCGGCGCGCTGCGGACCATCTGAGAGACTTAC 9049
8990 AGGACCGGCTGACCGCTGATGACAGGCGGCGCGCTGCGGACCATCTGAGAGACTTAC 9049
1441 CTTTGCAGATGCTGAGGCGGATGAGGCTTCACTGTGCGGACGAGCGCGCATGAGAG 1500
Db CTTTGCAGATGCTGAGGCGGATGAGGCTTCACTGTGCGGACGAGCGCGCATGAGAG 9109
9050 CTTTGCAGATGCTGAGGCGGATGAGGCTTCACTGTGCGGACGAGCGCGCATGAGAG 9109
1501 CACCGGCTGCGGAGGCGGCGGACCGTGGAGGAGGCGGCGGAGCGGCTGGAAGCTGCG 1560
Db CACCGGCTGCGGAGGCGGCGGACCGTGGAGGAGGCGGCGGAGCGGCTGGAAGCTGCG 9169
9110 CACCGGCTGCGGAGGCGGCGGACCGTGGAGGAGGCGGCGGAGCGGCTGGAAGCTGCG 9169
1561 GCGCAGGAGACAGACGTCGCCGCGGTGCGGAGTATGCGGATTCCTCAACGCGGCAAG 1620
Db GCGCAGGAGACAGACGTCGCCGCGGTGCGGAGTATGCGGATTCCTCAACGCGGCAAG 9229
9170 GCGCAGGAGACAGACGTCGCCGCGGTGCGGAGTATGCGGATTCCTCAACGCGGCAAG 9229
1621 CTCGCTTCTTCTTCAACGAGCAAGGCGGCGGACGCTGGGATGAGCGCTGAGCTTAC 1680
Db CTCGCTTCTTCTTCAACGAGCAAGGCGGCGGACGCTGGGATGAGCGCTGAGCTTAC 9289
9230 CTCGCTTCTTCTTCAACGAGCAAGGCGGCGGACGCTGGGATGAGCGCTGAGCTTAC 9289
1681 CATGATATGCTGCGGCTTCCGAGGCGGTTGCAACCTGTGCGGAGGCTGTTCAACGAGAG 1740
Db CATGATATGCTGCGGCTTCCGAGGCGGTTGCAACCTGTGCGGAGGCTGTTCAACGAGAG 9349
9290 CATGATATGCTGCGGCTTCCGAGGCGGTTGCAACCTGTGCGGAGGCTGTTCAACGAGAG 9349
1741 CTCGACCGGCGGCTCGGAGGAGTGTGAGCGGACCGGACGCGGCTGCAAGCGCGCTG 1800
Db CTCGACCGGCGGCTCGGAGGAGTGTGAGCGGACCGGACGCGGCTGCAAGCGCGCTG 9409
9350 CTCGACCGGCGGCTCGGAGGAGTGTGAGCGGACCGGACGCGGCTGCAAGCGCGCTG 9409
1861 CTCGACCGGCGGCTCGGAGGAGTGTGAGCGGACCGGACGCGGCTGCAAGCGCGCTG 1920
Db CTCGACCGGCGGCTCGGAGGAGTGTGAGCGGACCGGACGCGGCTGCAAGCGCGCTG 9529
9470 CTCGACCGGCGGCTCGGAGGAGTGTGAGCGGACCGGACGCGGCTGCAAGCGCGCTG 9529
1921 GTGCTGCTGCTGAGGCGGCGTGTCTGCTTGAAGAGCGGCTGCTGCTGCTGCTGCTG 1980
Db GTGCTGCTGCTGAGGCGGCGTGTCTGCTTGAAGAGCGGCTGCTGCTGCTGCTGCTG 9589
9530 GTGCTGCTGCTGAGGCGGCGTGTCTGCTTGAAGAGCGGCTGCTGCTGCTGCTGCTG 9589
1981 CGCGGCGCTGATGACAGCGCTGCGCGCGCGGCGGAGTGTGATTCGATGAGGCGCGG 2040
Db CGCGGCGCTGATGACAGCGCTGCGCGCGCGGCGGAGTGTGATTCGATGAGGCGCGG 9649
9590 CGCGGCGCTGATGACAGCGCTGCGCGCGCGGCGGAGTGTGATTCGATGAGGCGCGG 9649

QY	2041	GAGGCCGATGATGGCTGCTGGCGATGAGCCGCCACACCGTCTGGGTGTGATTCGCCCGGATC	2100
Dp	9650	GAGGCGCATGTGGTGTGGTGGGCGCCGCACACCGTCTGGGTGTGATTCGCCCGGATC	9709
QY	2101	AACGCTCCGAGACAGAGTGTGATGCGCGAGCGCCGGGAGAACCCGGTCATGTGATCGCGAGC	2160
Dp	9710	AACGCTCCGAGACAGAGTGTGATGCGCGAGCGCCGGGAGAACCCGGTCATGTGATCGCGAGC	9769
QY	2161	GCGATGACCGGCGCGGGGGCGCGAACCAAGCGCTCGACGTCTTGCAATGCGTTTCACTCA	2220
Dp	9770	GCGATGACCGGCGCGGGGGCGCGAACCAAGCGCTTCACGTCGTGCAATGCGTTTCACTCA	9829
QY	2221	CCGCTCATGTGCCCCGATGCTGTGAGAGCGTTGGGGGTGTGTGGCCGAGTGTGTACTACCG	2280
Dp	9830	CCGCTCATGTGCCCCGATGCTGTGAGAGCGTTGGGGGTGTGTGGCCGAGTGTGTACTACCG	9889
QY	2281	CGGCGCTCATGTGCTCTGTGTCAGCAATCTGAGCGGAGAGCTTGTCACACAGAGTGTGAGC	2340
Dp	9890	CGGCGCTCATGTGCTCTGTGTCAGCAATCTGAGCGGAGAGCTTGTCACACAGAGTGTGAGC	9949
QY	2341	TCGCGCGGGCTAATTGGGTGCGCCACGCGAGAGGTGTGTGCGCTTCGCGAGTGAAGTGAAG	2400
Dp	9950	TCGCGCGGGCTAATTGGGTGCGCCACGCGAGAGGTGTGTGCGCTTCGCGAGTGAAGTGAAG	10009
QY	2401	GCGCTGCAACGCGGCGCGGTGGCGGACCTTGTCAGAGTCTGTCCGAATGACGCTGTCTC	2460
Dp	10010	GCGCTGCAACGCGGCGCGGTGGCGGACCTTGTCAGAGTCTGTCCGAATGACGCTGTCTC	10069
QY	2461	GGCGCTGGAGCGCTGCTGTGTCATATGCGGAGAGCGCCGCGCGGCTGCTGTGCGTGGCGGCT	2520
Dp	10070	GGCGCTGGAGCGCTGCTGTGTCATATGCGGAGAGCGCCGCGCGGCTGCTGTGCGTGGCGGCT	10129
QY	2521	GGGCGGTGACGAGCCGCGCGACCGTGTCTGAGAGCGCTTCGCGCGGCTCTTGAGGCGCTGTGAGC	2580
Dp	10130	GGGCGGTGACGAGCCGCGCGACCGTGTCTGAGAGCGCTTCGCGCGGCTCTTGAGGCGCTGTGAGC	10189
QY	2581	CTGGTCTCTGTGGGCGCGCTCTTCTTCCCTCAGAGGGGGCGGCGGGTCCGCTGCCAGGTAC	2640
Dp	10190	CTGGTCTCTGTGGGCGCGCTCTTCTTCCCTCAGAGGGGGCGGCGGGTCCGCTGCCAGGTAC	10249
QY	2641	CCCTTGGGACGCGCGAGCGCTACTGGAATGACACAGAAACCGACAGCGCGGGCGGTGGGAGC	2700
Dp	10250	CCCTTGGGACGCGCGAGCGCTACTGGAATGACACAGAAACCGACAGCGCGGGCGGTGGGAGC	10309
QY	2701	CGCGGTGCTCCGAGAGCGGATCAAGACAGAGTTCAGAGAGGGGGCGCGGTGTCCGCGCGGC	2760
Dp	10310	CGCGGTGCTCCGAGAGCGGATCAAGACAGAGTTCAGAGAGGGGGCGCGGTGTCCGCGCGGC	10369
QY	2761	GACCGGCGGACGCTCGGCTTGACATCCGCGCGCCAGAGCGGACCGCCGAGAGAGGTCTC	2820
Dp	10370	GACCGGCGGACGCTCGGCTTGACATCCGCGCGCCAGAGCGGACCGCCGAGAGAGGTCTC	10429
QY	2821	GAGGCGCGCGCGGACCGTCCGTTCCGGGTGAGATTCGATAGCCAGGCGGTGTGGATTCAC	2880
Dp	10430	GAGGCGCGCGCGGACCGTCCGTTCCGGGTGAGATTCGATAGCCAGGCGGTGTGGATTCAC	10489
QY	2881	CTCGTGTCTTGGGTCACGAGCGGCGGCGCCCTGTGTCTTGGGCGAGGTTCAGATTCGCGTCTC	2940
Dp	10490	CTCGTGTCTTGGGTCACGAGCGGCGGCGCCCTGTGTCTTGGGCGAGGTTCAGATTCGCGTCTC	10549
QY	2941	GACGCGGCGGGGCTCAAGCTTCATATGATGTCCAGCTGCGCTGGGCGATGTGTCCCGAGAC	3000
Dp	10550	GACGCGGCGGGGCTCAAGCTTCATATGATGTCCAGCTGCGCTGGGCGATGTGTCCCGAGAC	10609
QY	3001	CTGCGCGGAGAAAGCCCAACCTTCGCGCTGTGTCTGGAAGCCAGTGTCCGCGGGCGCATGTCTC	3060
Dp	10610	CTGCGCGGAGAAAGCCCAACCTTCGCGCTGTGTCTGGAAGCCAGTGTCCGCGGGCGCATGTCTC	10669
QY	3061	GCCTGTGGGCGAGGCGGTGTAACGAGCTGTGGTGGGCGCAACCGGTTCATCCGCTTTTGGGCG	3120
Dp	10670	GCCTGTGGGCGAGGCGGTGTAACGAGCTGTGGTGGGCGCAACCGGTTCATCCGCTTTTGGGCG	10729
QY	3121	GGAGCGTTTGTCAACCACTCACCAAGTCCGAGTGGGTGGTGTGCTTCGAGCTCAAGGCG	3180

Db	10730	GGAGCGTTTGGTACCAAGTACACAGTCCGGCTGCGCTGGTCTGCTCTCGGCTTCAGAGCG	10789
Qy	3181	CTCTCGGCGATTCGAGGCGGCGGCGATGCGCTGCGCTAACGCGCATGGTACGCGCTC	3240
Db	10790	CTCTCGGCGATTCGAGGCGGCGGCGCATGCGCTGCGCTAACGCGCATGGTACGCGCTC	10841
Qy	3241	GACAGAAATAGCCGCTTTCAGCCGCGGGAGCGGGGTGTCTGATTCATGTCGCGTGACCGCGGG	3300
Db	10850	GACAGAAATAGCCGCTTTCAGCCGCGGGAGCGGGGTGTCTGATTCATGTCGCGTGACCGCGGG	10901
Qy	3301	GTCCGTCTCCCGCGGGGTGCGATGGGCGSACAGTGGGAGCCGAGGTCATATGGAGCGGCC	3360
Db	10910	GTCCGTCTCCCGCGGGGTGCGATGGGCGSACAGTGGGAGCCGAGGTCATATGGAGCGGCC	10966
Qy	3361	GGCACGCGCCGAGAAACGCGCCTTACCTTGAGTCCGCTCGGCGCTGCGGTAATGTAGCGATTCC	3420
Db	10970	GGCACGCGCCGAGAAACGCGCCTTACCTTGAGTCCGCTCGGCGCTGCGGTAATGTAGCGATTCC	11022
Qy	3421	CGCTCGGACCGGTTTCGTGCGCGACGTGCGCGCTGAGCGGCGCGAGGGAGTACGCTC	3480
Db	11030	CGCTCGGACCGGTTTCGTGCGCGACGTGCGCGCTGAGCGGCGCGAGGGAGTACGCTC	11088
Qy	3481	GTCTCTCAATCCGCTCTTCGGGCGAGCTGATTCGACAAAGTTTCAATCTCTCGCATATGCGAC	3540
Db	11090	GTCTCTCAATCCGCTCTTCGGGCGAGCTGATTCGACAAAGTTTCAATCTCTCGCATATGCGAC	11145
Qy	3541	GGCGCGTTTGTGAGAGCTCGGCGAAGCGGACCTGTTTACCGCGGATTAACAGCTCGGCGTTCGG	3600
Db	11150	GGCGCGTTTGTGAGAGCTCGGCGAAGCGGACCTGTTTACCGCGGATTAACAGCTCGGCGTTCGG	11207
Qy	3601	CCGTTCTCTGCGCAATCTCTCTCTCTCGTGTGTGATTCCTCGGGGGATGATGCTTCGACGG	3660
Db	11210	CCGTTCTCTGCGCAATCTCTCTCTCTCTCGTGTGTGATTCCTCGGGGGATGATGCTTCGACGG	11266
Qy	3661	CCGCGCGGGGTCCGTGCGCTCTTGTGAGAGAGCTCTCGGCGCTTATCCGCGGAGCGCTGTC	3720
Db	11270	CCGCGCGGGGTCCGTGCGCTCTTGTGAGAGAGCTCTCGGCGCTTATCCGCGGAGCGCTGTC	11321
Qy	3721	ACCCCTCCCCCATTCGCGAGCGCTCCGATGCGCCGTGTGCGCCGATGCGTTCGCGAGCATG	3780
Db	11330	ACCCCTCCCCCATTCGCGAGCGCTCCGATGCGCCGTGTGCGCCGATGCGTTCGCGAGCATG	11388
Qy	3781	GCGCAGGCGGACATCTTGGGAAAGCTCTGTACTACGCTGGGTGACCCCGAGGTTCCAGATC	3840
Db	11390	GCGCAGGCGGACATCTTGGGAAAGCTCTGTACTACGCTGGGTGACCCCGAGGTTCCAGATC	11445
Qy	3841	CGTATTTCCAAACCGACGAGGCGCGGCGCTCTCCACCGGGGATGCGGACCTGTCTGCACAGG	3900
Db	11450	CGTATTTCCAAACCGACGAGGCGCGGCGCTCTCCACCGGGGATGCGGACCTGTCTGCACAGG	11501
Qy	3901	CTGCGCGTCAAGTGTGCGCGGCGCGCGCGCGCGCGCGCTGAGGCGTTCCCTCGGTAACGACG	3960
Db	11510	CTGCGCGTCAAGTGTGCGCGGCGCGCGCGCGCGCGCGCGCTGAGGCGTTCCCTCGGTAACGACG	11566
Qy	3961	GTCCTCGCAGGTGTCTGCGACGCGCCGAAATCAGGTGCGCGGAGCGGCTGTTCAACCGCG	4020
Db	11570	GTCCTCGCAGGTGTCTGCGACGCGCCGAAATCAGGTGCGCGGAGCGGCTGTTCAACCGCG	11622
Qy	4021	CTCGGCAATGGAATCTGCTCATGCGCGGTGAGAGCTGCGCAATGTAATCGAGCGAGCTTCACG	4080
Db	11630	CTCGGCAATGGAATCTGCTCATGCGCGGTGAGAGCTGCGCAATGTAATCGAGCGAGCTTCACG	11688
Qy	4081	CTGAAGCTGTGACGAGAGTCTCTGTGCCAGTCCGCCCAATATGCGCTTGTGGGCGCAAAAC	4140
Db	11690	CTGAAGCTGTGACGAGAGTCTCTGTGCCAGTCCGCCCAATATGCGCTTGTGGGCGCAAAAC	11744
Qy	4141	CTGTGTGATGCTCTCGCCACAGGCTCTCTCTTGTGAGCGGAGTGGCGCGCGAGAACTTACGG	4200
Db	11750	CTGTGTGATGCTCTCGCCACAGGCTCTCTCTTGTGAGCGGAGTGGCGCGCGAGAACTTACGG	11807
Qy	4201	GCAGGCGGTGCAAAACGATCTGTCTCATTCGGGCGCAGATCAAGACTTGGGAAATCAATTGCC	4260

Db 11810 GCAGCGGTGCAGAAAGCACTTGTCTCATCGGGGCGAGATCAAGACTGGAAATCATTTGCC 11869
Qy 4261 CTATGA 4266
Db 11870 CTATGA 11875

RESULT 7
AR208671
LOCUS AR208671 68750 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6383787.
ACCESSION AR208671
VERSION AR208671.1 GI:21509886
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 68750)
AUTHORS Schupp,T., Ligon,J.,Madison., Molnar,I., Zirkle,R., Cyr,D.,Dawn., and
Goriach,J.
TITLE Genes for the biosynthesis of epothilones
JOURNAL Patent: US 6383787-A 1 07-MAY-2002;
FEATURES
source 1..68750
/organism="Unknown"
BASE COUNT 9596 a 22456 c 25539 g 11159 t
ORIGIN

Query Match 100.0%; Score 4266; DB 6; Length 68750;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGGGGGATTCGTCCTCATCGAGCGCGCAAGATCCGATTGCCATCGTGGAGCGAGT 60
Db 7610 GTGGGGGATTCGTCCTCATCGAGCGCGCAAGATCCGATTGCCATCGTGGAGCGAGT 7669
Qy 61 TGCCGCTGCGCCGATGCGTGCATCGATCGAGCGGGTCTGAGCGCTCTCGAGGAGCTG 120
Db 7670 TGCCGCTGCGCCGATGCGTGCATCGATCGAGCGGGTCTGAGCGCTCTCGAGGAGCTG 7729
Qy 121 CGCGACACCGTGGGCGAGTCCCGCGCAACGCTGGGATCGACGACGCTGTTGATCCC 180
Db 7730 CGCGACACCGTGGGCGAGTCCCGCGCAACGCTGGGATCGACGACGCTGTTGATCCC 7789
Qy 181 GACCCCGATGCCCCGGGGAAGACGCCCTTACGCCGCGCATCTTTCTTGAGCGATAGCC 240
Db 7790 GACCCCGATGCCCCGGGGAAGACGCCCTTACGCCGCGCATCTTTCTTGAGCGATAGCC 7849
Qy 241 TGCTTCGAGCGCTCTTCTTTCGCACTCTGCGCTCGGAAAGCGCTGGAGTGAACCTGCA 300
Db 7850 TGCTTCGAGCGCTCTTCTTTCGCACTCTGCGCTCGGAAAGCGCTGGAGTGAACCTGCA 7909
Qy 301 CATCGACTTGTCTGAGAGGTGTGTGGAGGCGCTGGAGAACGCGCGCATGCTCCATCG 360
Db 7910 CATCGACTTGTCTGAGAGGTGTGTGGAGGCGCTGGAGAACGCGCGCATGCTCCATCG 7969
Qy 361 GGGCTGCTGCGGATCGGAAACGGGAGTTCATCGGGATGCGCCGCTCGGAATATAGAGCC 420
Db 7970 GGGCTGCTGCGGATCGGAAACGGGAGTTCATCGGGATGCGCCGCTCGGAATATAGAGCC 8029
Qy 421 GCGCTGCGCAAGCAGCGCGCTCCGACAGATCGAGCTCATGCGGCGTGGGAGAGATG 480
Db 8030 GCGCTGCGCAAGCAGCGCGCTCCGACAGATCGAGCTCATGCGGCGTGGGAGAGATG 8089
Qy 481 CCAGCGTTCGAGCGGCGGCAATCTGTATGCTTCGGGCTCGAGAGGCGGTGTGCGG 540
Db 8090 CCAGCGTTCGAGCGGCGGCAATCTGTATGCTTCGGGCTCGAGAGGCGGTGTGCGG 8149
Qy 541 GTGGATACGGCTTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
Db 8150 GTGGATACGGCTTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8209
Qy 601 TCCGGGGAATGCTCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

Db 8210 TCCGGGGAATGCTCAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8269
Qy 661 CTGCTGTGCTCTCGAAGACCCCGGCGCTGCGCCAGAGAGCGGTCTGTCAGAGCATTTTCG 720
Db 8270 CTGCTGTGCTCTCGAAGACCCCGGCGCTGCGCCAGAGAGCGGTCTGTCAGAGCATTTTCG 8329
Qy 721 GCGAGGCGCATGCGTTGGATCGAGCGAAGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db 8330 GCGAGGCGCATGCGTTGGATCGAGCGAAGGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8389
Qy 781 AGTGAAGCCCGCGCGAGAGCGAGATCGATATTTGGCGGATTCAGATATCCGCAATCAT 840
Db 8390 AGTGAAGCCCGCGCGAGAGCGAGATCGATATTTGGCGGATTCAGATATCCGCAATCAT 8449
Qy 841 CACGACGCTGCGAGCAGCGGTCTGACCGTGCAGAACGCGAGCTCCCAAGAAATGCTGCTG 900
Db 8450 CACGACGCTGCGAGCAGCGGTCTGACCGTGCAGAACGCGAGCTCCCAAGAAATGCTGCTG 8509
Qy 901 AAACGGGCGCTGGGAGAGCGAGCGTGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
Db 8510 AAACGGGCGCTGGGAGAGCGAGCGTGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8569
Qy 961 GGCACGGGCGAGCGCTTGTGACCCCATCGAAATCCAAAGCTGGAATGCGGTATACCGC 1020
Db 8570 GGCACGGGCGAGCGCTTGTGACCCCATCGAAATCCAAAGCTGGAATGCGGTATACCGC 8629
Qy 1021 CTGCGGCGAGATGTCGCCACCGCGCTGCTGATCGGCTGCTGGAAGACCACTTGGCCAT 1080
Db 8630 CTGCGGCGAGATGTCGCCACCGCGCTGCTGATCGGCTGCTGGAAGACCACTTGGCCAT 8689
Qy 1081 CTTGATGATGCTGGGGGATCACTGCGGCTGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
Db 8690 CTTGATGATGCTGGGGGATCACTGCGGCTGGAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8749
Qy 1141 CAGATTCTGCGCACTTCACGCGCAGCGCTGGAACCCCGGATCTCATGCGGCTGATCTT 1200
Db 8750 CAGATTCTGCGCACTTCACGCGCAGCGCTGGAACCCCGGATCTCATGCGGCTGATCTT 8809
Qy 1201 CGGCTGACCTGACCGCGCGCGCGGACACCGTGGCGCGACTGGAATACCGCGCGAGCGCG 1260
Db 8810 CGGCTGACCTGACCGCGCGCGCGGACACCGTGGCGCGACTGGAATACCGCGCGAGCGCG 8869
Qy 1261 GGGGTGAGCTGCTGCGGATGAGCGGAGCAACCGCGACGCGTGGTGGAAAGAGCGCGCG 1320
Db 8870 GGGGTGAGCTGCTGCGGATGAGCGGAGCAACCGCGACGCGTGGTGGAAAGAGCGCGCG 8929
Qy 1321 GCGGCGAGTGCACACCGCGCGCGCGGAGCGACCGGAGAGCTGCTGCTGCGCA 1380
Db 8930 GCGGCGAGTGCACACCGCGCGCGCGGAGCGACCGGAGAGCTGCTGCTGCGCA 8989
Qy 1381 AGAACCGGCTGACGCTTGGATGACAGCGCGCGCGGCTGCGGACATCTGGAACCTTAC 1440
Db 8990 AGAACCGGCTGACGCTTGGATGACAGCGCGCGCGGCTGCGGACATCTGGAACCTTAC 9049
Qy 1441 CTTTGCAGATGCTGCGGAGATGAGCGCTTCACTGCGGAGCAAGCGCGAGCGCGATGAG 1500
Db 9050 CTTTGCAGATGCTGCGGAGATGAGCGCTTCACTGCGGAGCAAGCGCGAGCGCGATGAG 9109
Qy 1501 CACCGGCTCGCGGTGCGCGAGCTGCAAGGAGGCTGCGGCGAGCGCTGCGACGCTGCG 1560
Db 9110 CACCGGCTCGCGGTGCGCGAGCTGCAAGGAGGCTGCGGCGAGCGCTGCGACGCTGCG 9169
Qy 1561 GCGCAGGAGCAGAGCTGCGCGCGGTGCGGCTGCGGAGATTCGCGCATCTTCAACGCGGCAAG 1620
Db 9170 GCGCAGGAGCAGAGCTGCGCGCGGTGCGGCTGCGGAGATTCGCGCATCTTCAACGCGGCAAG 9229
Qy 1621 CTGCGCTTCTTCTTCAACGAGCAAGGCGGAGAGCGTGGGATGAGCGCGTGGGCTGTAC 1680
Db 9230 CTGCGCTTCTTCTTCAACGAGCAAGGCGGAGAGCGTGGGATGAGCGCGTGGGCTGTAC 9289
Qy 1681 GATGATGATCGCGCTTCCGAGAGCGTTCACCTGTGCTGAGGCTGTTCAACGAGAG 1740

Dh 9290 GATGTAATGTCCTCGGCTTCGCGAGGCGTTTCAGACTGTGCTGAGGCTGTTCACACGAGAG 9349
Qy 1741 CTGACACCGGCGCTCCGCGAGGTGATGTGGCCGGAACCGGCCAGGCTGAGACCGCGCTG 1800
Dh 9350 CTGACACCGGCGCTCCGCGAGGTGATGTGGCCGGAACCGGCCAGGCTGAGACCGCGCTG 9409
Qy 1801 CTGACACGAGACCTTCAACCCAGCGCGCTGTTCACCTTCGAATATGCGCTCGCGCG 1860
Dh 9410 CTCACACGAGACCTTCAACCCAGCGCGCTGTTCACCTTCGAATATGCGCTCGCGCG 9469
Qy 1861 CTGTGGCGGTGTGGGGGTGTAGAGCGGAGTGTGTGCGCGGCATATGATGATGATGATG 1920
Dh 9470 CTGTGGCGGTGTGGGGGTGTAGAGCGGAGTGTGTGCGCGGCATATGATGATGATGATG 9529
Qy 1921 GTGGCTGCTGCTGATGCGGCGGCTGTTCCTGCTTGAAGACGCGGTTCCTGATGATGATG 1980
Dh 9530 GTGGCTGCTGCTGATGCGGCGGCTGTTCCTGCTTGAAGACGCGGTTCCTGATGATGATG 9589
Qy 1981 CGCGGCGCGCTGATGCGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
Dh 9590 CGCGGCGCGCTGATGCGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9649
Qy 2041 GAGGCGCGATGTGGCTGCTGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
Dh 9650 GAGGCGCGATGTGGCTGCTGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9709
Qy 2101 AACGCTCCGAGACAGAGTGTATCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2160
Dh 9710 AACGCTCCGAGACAGAGTGTATCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9769
Qy 2161 GCGATGCG 2220
Dh 9770 GCGATGCG 9829
Qy 2221 CCGGCTCATGCG 2280
Dh 9830 CCGGCTCATGCG 9889
Qy 2281 CCGGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
Dh 9890 CCGGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 9949
Qy 2341 TCGGCGGCTATTTAGGTGCGCCAGCGCGGAGAGGTGTGCGCTTCGCGGATGAGTGAAG 2400
Dh 9950 TCGGCGGCTATTTAGGTGCGCCAGCGCGGAGAGGTGTGCGCTTCGCGGATGAGTGAAG 10009
Qy 2401 GCGCTGCAACG 2460
Dh 10010 GCGCTGCAACG 10069
Qy 2461 GCGCTGCAACG 2520
Dh 10070 GCGCTGCAACG 10129
Qy 2521 GGGCGGTGACGAGCG 2580
Dh 10130 GGGCGGTGACGAGCG 10189
Qy 2581 CTGTGCTGCTGCGGCGCGCTTCTTCCCTCAGGCGGCGCGCGCGCGCGCGCGCGCG 2640
Dh 10190 CTGTGCTGCTGCGGCGCGCTTCTTCCCTCAGGCGGCGCGCGCGCGCGCGCGCG 10249
Qy 2641 CTTTGGGAGCGGAGCGCTTACCTGATGACAGAAAGCGGAGCGAGCGCGCGCGCGCG 2700
Dh 10250 CTTTGGGAGCGGAGCGCTTACCTGATGACAGAAAGCGGAGCGAGCGCGCGCGCGCG 10309
Qy 2701 CGCGGTGCTCGGAGCGGAGCGAGCGAGGTGAGAGAGGAGGCGCGCGCGCGCGCGCG 2760
Dh 10310 CGCGGTGCTCGGAGCGGAGCGAGCGAGGTGAGAGAGGAGGCGCGCGCGCGCGCG 10369
Qy 2761 GACCGGCGGAGCGCTGCGCTGAGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2820
Dh 10370 GACCGGCGGAGCGCTGCGCTGAGACATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10429

Qy 2821 GAGCG 2880
Dh 10430 GAGCG 10489
Qy 2881 CTGTGCTGCTGCGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2940
Dh 10490 CTGTGCTGCTGCGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10549
Qy 2941 GAGCG 3000
Dh 10550 GAGCG 10609
Qy 3001 CTGCGGAGGAAAGCG 3060
Dh 10610 CTGCGGAGGAAAGCG 10669
Qy 3061 GCGGTGCGCGAGGCGGTGAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3120
Dh 10670 GCGGTGCGCGAGGCGGTGAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 10729
Qy 3121 GGAAGCTTGTGTACCAAGTCAACAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180
Dh 10730 GGAAGCTTGTGTACCAAGTCAACAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 10789
Qy 3181 CTTCGCGGATCGAGGCG 3240
Dh 10790 CTTCGCGGATCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10849
Qy 3241 GACAGAAATAGCGCGCTTCAGCGCGGAGAGCGGTGTGATCATGCGGCGAGCGCGCG 3300
Dh 10850 GACAGAAATAGCGCGCTTCAGCGCGGAGAGCGGTGTGATCATGCGGCGAGCGCG 10909
Qy 3301 GTGCGTCTGCG 3360
Dh 10910 GTGCGTCTGCG 10969
Qy 3361 GGCAGCGCGCGGAAAGCGCGCTTACCTGAGAGTGTGCGGCGCGCGCGCGCGCGCG 3420
Dh 10970 GGCAGCGCGCGGAAAGCGCGCTTACCTGAGAGTGTGCGGCGCGCGCGCGCGCG 11029
Qy 3421 CGCTCGAGCGGCTTGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3480
Dh 11030 CGCTCGAGCGGCTTGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11089
Qy 3481 GTGCTCAACTGCGCTGCGCGCGAGCTGATGAGCAAGATTTCATCTCTGCGCATCG 3540
Dh 11090 GTGCTCAACTGCGCTGCGCGAGAGCTGATGAGCAAGATTTCATCTCTGCGCATC 11149
Qy 3541 GGCAGGTTTGTGAGGCTCGGAGAGCGGAGCTGTTACCGGAGTAACAGAGCTCGGCG 3600
Dh 11150 GGCAGGTTTGTGAGGCTCGGAGAGCGGAGCTGTTACCGGAGTAACAGAGCTCGG 11209
Qy 3601 CCGTTCCTGCGGATCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3660
Dh 11210 CCGTTCCTGCGGATCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11269
Qy 3661 CCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3720
Dh 11270 CCGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 11329
Qy 3721 ACCCTTCCCGCATCGGAGCGCTCCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3780
Dh 11330 ACCCTTCCCGCATCGGAGCGCTCCGATGCGCGCGCGCGCGCGCGCGCGCGCGCG 11389
Qy 3781 GCGAGGCGGAGCATCTTGTGAGAGCTGTACTACGCTGTGAGAGCGCGGAGTCCAGATC 3840
Dh 11390 GCGAGGCGGAGCATCTTGTGAGAGCTGTACTACGCTGTGAGAGCGCGGAGTCCAG 11449
Qy 3841 CGTATTCCAAACCAACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3900
Dh 11450 CGTATTCCAAACCAACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11509

Oy	3901	CTCCGCGTCAGCTGGCGCGCGCGCGCGCGCGCTGGAGGCGTTCTCCGTAACGAG	3960
Db	11510	CTCCGCGTCAGCTGGCGCGCGCGCGCGCGCGCTGGAGGCGTTCTCCGTAACGAG	11566
Oy	3961	GTCCTCGACAGTGTCTCGCGCACGCGCCGGAATCAAGGTCCGCGCGAGGCGCTTTCAACCGC	4020
Db	11570	GTCCTCGACAGGTCGTCCGCGACGCGCGGAATCAAGGTCCGCGCGAGGCGCTTTCAACCGC	11622
Oy	4021	CTCCGCGATGACCTCGCTCTCATATGAGCCGTGGAGCTGGCGCAATGGTATCCAGGGGAGCTTCAG	4080
Db	11630	CTCCGCGATGACCTCGCTCTCATATGAGCCGTGGAGCTGGCGCAATGGTATCCAGGGGAGCTTCAG	11688
Oy	4081	CTGAAGCTGTGACGACGAGTCTCTGTCCACGCTCCCCCAATATTCGCTTATGGCCCAAAAC	4140
Db	11690	CTGAAGCTGTGACGACGAGTCTCTGTCCACGCTCCCCCAATATTCGCTTATGGCCCAAAAC	11744
Oy	4141	CTGTTGATGTCCTTCGCGCACAGCTCTCTCTTTGAGGCGGATGGCGCGCGAGAACTTACGG	4200
Db	11750	CTGTTGATGTCCTTCGCGCACAGCTCTCTCTTTGAGGCGGATGGCGCGCGAGAACTTACGG	11800
Oy	4201	GCAGGCGGTGCAAAACGACCTTCTCATTCGGGCGCAGATCAGACTGGGAAATCATTTGCC	4260
Db	11810	GCAGGCGGTGCAAAACGACCTTCTCATTCGGGCGCAGATCAGACTGGGAAATCATTTGCC	11866
Oy	4261	CTATGCA 4266	
Db	11870	CTATGCA 11875	
RESULT 8			
AF217189			
LOCUS	AF217189	58733 bp	linear BCT 09-JUN-2000
DEFINITION	Sorangium cellulosum putative transposase gene, partial cds;		
	putative transposase gene, complete cds; epothilone biosynthesis		
	gene cluster; complete sequence; putative membrane protein gene,		
	complete cds.		
ACCESSION	AF217189		
VERSION	AF217189.1	GI:7453554	
KEYWORDS			
SOURCE	Polyangium cellulosum		
ORGANISM	Polyangium cellulosum		
	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;		
	Soranginae; Polyangiaceae; Polyangium.		
REFERENCE	1 (bases 1 to 58733)		
AUTHORS	Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and		
JULIEN,B.			
TITLE	Cloning and heterologous expression of the epothilone gene cluster		
JOURNAL	Science 287 (5453), 640-642 (2000)		
MEDLINE	20115953		
PUBMED	10649995		
REFERENCE	2 (bases 1 to 58733)		
AUTHORS	Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.		
TITLE	Isolation and characterization of the epothilone biosynthetic gene		
JOURNAL	cluster from Sorangium cellulosum		
MEDLINE	Gene 249 (1-2), 153-160 (2000)		
PUBMED	20293058		
REFERENCE	3 (bases 1 to 58733)		
AUTHORS	Julien,B.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center		
	Place, Hayward, CA 94545, USA		
FEATURES	Location/Qualifiers		
source	1..58733		
	/organism="Polyangium cellulosum"		
	/mol_type="genomic DNA"		
	/strain="SNP4"		
	/db_xref="taxon:56"		
	<1..992		
	/note="orfA"		
	/codon_start=3		
	/transl_table=11		
	/product="putative transposase"		

gene
CDS
/protein_id="AAFG2878.1"
/db_xref="GI:7453555"
/translation="VRGVHEVAFDGGVPRVLTVDNKNMAYNVERGDAIRPHPTLLAL
SADYFERPRPAVAFDGERSVLLRHPDTPRDEHERVDEGKTPVARPLDNTSVPH
DTRTRTLVADLSQVRIADENQIVATVHWSMGQDQIQEOPHQRIVDEKREHRT
GLDIRARARSSQAFRLRVARGDGVSALIRLLQIDVACALAEALVEVLERDTRI
HIGVROVIDRREBRHLPVPSIVPTRGHVALVTPHSLTVDALKKDPTP"
989..1501
/note="orfB"
/codon_start=1
/transl_table=1
/product="putative transposase"
/protein_id="AAFG6879.1"
/db_xref="GI:7453556"
/translation="MTDITPTETKRLKSLGFLCALQEQBLADKPKMLREVLIAEER
RHKSLERLXNSRVAEFKPTMTDPSMPPKIDIEAVDDLYDSRYADLVEVTRVD
AOKPLISTNKAFAFMGCVFPHAACVTLVDRLVHRVAVIEIEBASVRLKAEKELNAN
RTKORRTKQH"
1998..6263
/gene="epoA"
1998..6263
/gene="epoA"
/note="type I polyketide synthase; epoH1 lone PKS loading
module"
/codon_start=1
/transl_table=1
/product="EpoA"
/protein_id="AAFG6280.1"
/db_xref="GI:7453557"
/translation="MADRPITRAEDPIAIVAGCRLPGGVIDLSGFWTLLGSDVTV
GVVAPERDAAVAFDPLDAPGKTPTVRASEPLSNAPDASFKPIISRELRMDPARH
LLERCALEALNIAIPBALVGTGTGVTIGTSPEYRALPRATASIEIDHSGLTGM
PSVAGRIISYVLGRGPCVADVTVYSSSLVAVHLACQSLSGECSTALAGVSIMLSB
STLWASTRLATDGRCAVDSADFGREGCAVVVLRKLSGARADGRIILAVIRG
SAIHNDGASGLTVPNGSSOEIVLKRALADAGCAASGVYEAHGTGGLDPIEIOAN
LNAVYGRDVAFTLLIGSVKTNLGHPEVASGITLKVLSLQHGOIPALHQAOLAN
PRISMGDIRLVYTAARTPWPMNTPPRAGVSSRMSGTMAHVTEEAAPATCPAPR
RPAELVVSATTAALADHARLDHLETTIPSOCLGVATSLATTSAMERLVAATN
SSGEFRLALDAAQGGQTPPEVVGRIIASSRKLAFLETPGAQQLTGMGRGLYDVPAR
REAPFLCRALFNOELDPLREVMVAPASVDALIDOTAFPQPLPFEELVAALMSS
MGVEBELVAGSISGELVAAVAGVSLIEDAFLVLAAGRLAPCAGMVSIAAPSA
DVAAVVAHVASISIAVNGDQVVIAGQPPVAIAAAMARCAKALVSHAFHS
PLMAPLEAFERVAESVSYRPSIVLYSNLSGKGTDEVSFGVWHAHEVVPADG
PLMAHAAAGFVEVGEKSTLLGLVPACLPARPALLASSAGDEPATVLEALGML
AVGLVAVSAGLFPESGRRVLPPTYPMQERERYIDTKADDAARGORAPAGCHDEVEK
GAVGGDSRRSRLDHPPESEGRREKVEAAGRPRLEIDEGVLDRLVLRTERRAGC
LGEVIEVDAGLSFNDVQLAGVPPDLPEKPNPPLLGGECAGRIYAEGVNGVH
GCVPIALISACAFAHTVTTSAOVLPRPQALSAEAMAPAVATVAVTAVLDGIRLQ
GERTVLIHAATGCGVLAIVQAHVGAETHATGPEKRAVLESIGVYVDSRDRPV
ADVRAWITGEGEVDVNLSELIDKSFNLRSKRVELEKRCVADNOLGSPFLR
NLSBVLDRMMLERPARVAFEEELGILAAGVFTPPPIATLPVARVADAFRSMQ
AOHKGILVLTGDEVOVQRIPTHAGAGPSTGRDRLRLRSAPAPAAALAEALRFO
VSOVARTPEIVGAELFTRLGMSLMAVELRNRIEASLKLSKSTFTLSPTNALTT
ONLIDALATALSERVAENLRAGVQSDPVSAGADQMEIALL"
6260..10492
/gene="epos"
6260..10492
/gene="epos"
/note="non ribosomal peptide synthetase; epoH1 lone PKS
module 1"
/codon_start=1
/transl_table=1
/product="EpoB"
/protein_id="AAFG6281.1"
/db_xref="GI:7453558"
/translation="WTINQLNLEHOGVTKLADGERLQIDAPKALNPNTLAISEH
KSTLLTMLRQLPAESIVPAERHVPPEPLTDIGSTWLGTAFTVPSGHAYRETD
CTDLDVAALSRFKVVARHMDLRAHTLPDMQOIVEPVADNIEIDILRGDSRTREA
RLVSLADMSRIYDTERPPPLHYVAAVLDQOQRLVYSIDLINLDGSGSLIKEDWY
SFPEEPERSLVEISRYDYLAEKRSKSAHORSADWGRVAEVLPPMLPMKAD
PSTLREIRFRTTEMLPESDSWRLKQVGEGLPTGVIIAASEVILGWSASPRFLD
NTLLENRLVPEPRVNDITGDTSNVLDIDITTRKSEFGQAKRIQEOQLMEMHCDVVS

QY	841	CACGACGGGTGCGAGCAGCGGGTCTTGACCGTGGCCGAACGGGAGCTCCCAAGAAATCGTCTG	900
Db	2838	CACGACGGAGCGAGCAGCGGGTCTTGACCGTGGCCGAACGGGAGCTCCCAAGAAATCGTCTG	2897
QY	901	AAACGGGGCCCTGGCGGGAGCAGCAGGCGTGGCCGCGCTCTTGCGTGGGGTTATAGTGAGGCAAC	960
Db	2898	AAACGGGGCCCTGGCGGGAGCAGCAGGCGTGGCCGCGCTCTTGCGTGGGGTTATAGTGAGGCAAC	2955
QY	961	GGCACGGGCACGACCGCTTGGTGAACCCATCGAAATCGAAGCTTGAAATGCGGTATACGGC	1020
Db	2958	GGCACGGGCACGACCGCTTGGTGAACCCATCGAAATCGAAGCTTGAAATGCGGTATACGGC	3017
QY	1021	CTCGGGCGAGATGTGCCACCGCCGCTGCTGATCGGGTGGGTGAAGACCAACTTGGCCAT	1088
Db	3018	CTCGGGCGAGATGTGCCACCGCCGCTGCTGATCGGGTGGGTGAAGACCAACTTGGCCAT	3077
QY	1081	CTGAGTATGGGTGGGGGTATCATCTGGGGCTGAGAGTGTCTTCCCTTCAGCAGGG	1144
Db	3078	CTGAGTATGGGTGGGGGTATCATCTGGGGCTGAGAGTGTCTTCCCTTCAGCAGGG	3137
QY	1141	CAGATTCTCTGGCGCACTTCCACGCGCGAGGCGCTGAAACCCCGGATCTCATGGGGTGAATCTT	1200
Db	3138	CAGATTCTCTGGCGCACTTCCACGCGCGAGGCGCTGAAACCCCGGATCTCATGGGGTGAATCTT	3197
QY	1201	CGGCTGACCGGTACCGCGCGCCCGGACACCGTGGCCGGAAGTGAATACCGCGGACGGCG	1260
Db	3198	CGGCTGACCGGTACCGCGCGCCCGGACACCGTGGCCGGAAGTGAATACCGCGGACGGCG	3257
QY	1261	GGGGGTGAGCTGGTTGGGATGAGCGGGGACCAACGGCGCATGTGGTCTGTGAAGAAGCGCG	1320
Db	3258	GGGGGTGAGCTGGTTGGGATGAGCGGGGACCAACGGCGCATGTGGTCTGTGAAGAAGCGCG	3317
QY	1321	GCGCGGACGTGCACACCGCGCGCGCGGAGCGACCGGACAGACTGCTGTCTGCACA	1380
Db	3318	GCGCGGACGTGCACACCGCGCGCGCGGAGCGGAGCGGACAGACTGCTGTCTGTCTGCACA	3377
QY	1381	AGGACCGCGGTACAGCCCTGGATGTCAACAGCGCGCGCGCTGGCGACATCTGGAGCTTAC	1440
Db	3378	AGGACCGCGGTACAGCCCTGGATGTCAACAGCGCGCGCGCTGGCGACATCTGGAGCTTAC	3437
QY	1441	CGTTGGCAGTGTCTGGGGGATGTGGCGTTCACTGAGGAGACGACGGCGAGGCGCATGGAG	1500
Db	3438	CGTTGGCAGTGTCTGGGGGATGTGGCGTTCACTGAGGAGACGACGGCGAGGCGCATGGAG	3497
QY	1501	CACCGGCTCGCGGGTGGCGGCACTCTGAGAGGAGGCTGGCGGACCGCTGGACGCTCTCG	1560
Db	3498	CACCGGCTCGCGGGTGGCGGCACTCTGAGAGGAGGCTGGCGGACCGCTGGACGCTCTCG	3555
QY	1561	GCGCAGGGACACAGCCTCGCGCGGGTGGCGGCACTATGCCGATTTCTCACGCGGCAAG	1620
Db	3558	GCGCAGGGACACAGCCTCGCGCGGGTGGCGGCACTATGCCGATTTCTCACGCGGCAAG	3617
QY	1621	CTCGCGCTTCTCTTACACGGGACAGGGGGCGGACGCTGGGCGATGGGCGGTGGCTGATC	1680
Db	3618	CTCGCGCTTCTCTTACACGGGACAGGGGGCGGACGCTGGGCGATGGGCGGTGGCTGATC	3677
QY	1681	GATGTATGTCGCGCGCTTCGCGGAGGCGCTTGCACCTGTGCGTGAAGGCTGTTCAACCAAGAG	1740
Db	3678	GATGTATGTCGCGCGCTTCGCGGAGGCGCTTGCACCTGTGCGTGAAGGCTGTTCAACCAAGAG	3737
QY	1741	CTCGACCGGCGCGCTTCGCGAGGTGATGTGGGCCGAACCGGCCAGCTGACGCGCGCTG	1800
Db	3738	CTCGACCGGCGCGCTTCGCGAGGTGATGTGGGCCGAACCGGCCAGCTGACGCGCGCTG	3797
QY	1801	CTCGACCGACAGCGCTTGCACCGACGGCGCGCTGTTCACCTTCGATATAGCGTGGCGCGG	1866
Db	3798	CTCGACCGACAGCGCTTGCACCGACGGCGCGCTGTTCACCTTCGATATAGCGTGTCCCGCG	3857
QY	1861	CTGTGGCGGTGCTGGGGGTGTAGAGCGCGGAGTTGTGCGCGGACATGAGCATCGTGAAGCTG	1920
Db	3858	CTGTGGCGGTGCTGGGGGTGTAGAGCGCGGAGTTGTGCGCGGACATGAGCATCGTGAAGCTG	3917
QY	1921	GTGGCTGCTGCGTGGCGGGCGGTGTTCTTCGCTTAGAGACCGCGGTGTTCTGTGGTGGCTGCG	1980

Db	3918	GTGGCTGCTGCGTGGGGGCGGTCTTCGCTTGAAGACGGCGGTCTCTGTGGCTGCG	3977
Qy	1981	CGCGGCGCCCTGATGCAAGCGCTTCGCGCGCGCGCGGCGATGTGTGATTCAGACGCG	2040
Db	3978	CGCGGGCGCCGATGCAAGCGCTTCGCGCGCGCGCGGCGATGTGTGATTCGCGCGCG	4037
Qy	2041	GAGCGCGATGTGTGCTGTGCGGTGGCGCGCAGCGACGCTGTGTGATTCGCGCGCTC	2100
Db	4038	GAGCGCGATGTGTGCTGTGCGGTGGCGCGCAGCGACGCTGTGTGATTCGCGCGCTC	4097
Qy	2101	AACCGTCCGACCAAGGTGTATGCGGGGGCGGGGAAACCGGTGATGTGATGCGGGC	2160
Db	4098	AACCGTCCGACCAAGGTGTATGCGGGGGCGGGGAAACCGGTGATGTGATGCGGGC	4157
Qy	2161	GCGATGGCCGCGCGCGGGGCGCGAACCAAGCGCTTCACGCTTCGCAATGCGTTCACTCA	2220
Db	4158	GCGATGGCCGCGCGCGGGGCGCGAACCAAGCGCTTCACGCTTCGCAATGCGTTCACTCA	4217
Qy	2221	CGCGTCATGGCCCCGATGTGAGGCGTTGGGCGTGTGGCCGACGTGCTGAGCTTACCG	2280
Db	4218	CGCGTCATGGCCCCGATGTGAGGCGTTGGGCGTGTGGCCGACGTGCTGAGCTTACCG	4277
Qy	2281	CGCGCTTCGATTCGTCTGTGTACGAATCTGACGGGAAGGCTTGCACAGACGAGGTAGC	2340
Db	4278	CGCGCTTCGATTCGTCTGTGTACGAATCTGACGGGAAGGCTTGCACAGACGAGGTAGC	4337
Qy	2341	TCGCGGGGCTATTGGGTGCGCCAGCGCGCAGAGGTGTGGCTTTCGCGATGTGAGTAG	2400
Db	4338	TCGCGGGGCTATTGGGTGCGCCAGCGCGCAGAGGTGTGGCTTTCGCGATGTGAGTAG	4397
Qy	2401	CGCGTCACGCGCGCGGTGGGGGCACTTGTGAGGTGGTCCGAATGACGCTGCTC	2460
Db	4398	CGCGTCACGCGCGCGGTGGGGGCACTTGTGAGGTGGTCCGAATGACGCTGCTC	4457
Qy	2461	GCGCTGATGGCTGTGATGCGCGGAGCGCGCGCGCGGTGCTCGCATGTGCGCGGCT	2520
Db	4458	GCGCTGATGGCTGTGATGCGCGGAGCGCGCGCGCGGTGCTCGCATGTGCGCGGCT	4517
Qy	2521	GGGCGTGACGAGCGCGGCGACCGTGTCTGAGGCGCTTCGCGCGGCTTCTGGGCGTGTGGC	2580
Db	4518	GGGCGTGACGAGCGCGGCGACCGTGTCTGAGGCGCTTCGCGCGGCTTCTGGGCGTGTGGC	4577
Qy	2581	CTGGTCTCTGGGCGCGCTTCTTCCCTCAGGGGGGCGGGGCTGCCGTGCCACGTAC	2640
Db	4578	CTGGTCTCTGGGCGCGCTTCTTCCCTCAGGGGGGCGGGGCTGCCGTGCCACGTAC	4637
Qy	2641	CTTTGGCAGCGCGACCGCTACTGGAATGACAAGCAACCGACACGCGGGGCGTGGGAC	2700
Db	4638	CTTTGGCAGCGCGACCGCTACTGGAATGACAAGCAACCGACACGCGGGGCGTGGGAC	4697
Qy	2701	CGCGGTGCTCGGAGGCGGCTCAGACGAGGTGACAGAGGGGGGCGCGGTGCGCGCGC	2760
Db	4698	CGCGGTGCTCGGAGGCGGCTCAGACGAGGTGACAGAGGGGGGCGCGGTGCGCGCGC	4757
Qy	2761	GACCGGCGACGCTCGGCTGACCATTCGCGCGCCGACAGCGGACCCCGGAGAGGTC	2820
Db	4758	GACCGGCGACGCTCGGCTGACCATTCGCGCGCCGACAGCGGACCCCGGAGAGGTC	4817
Qy	2821	GAGCGCGCGCGGACCGTCCGTTCCGGCTGAGATGATAGCCAGGCGGTGATGACAC	2880
Db	4818	GAGCGCGCGCGGACCGTCCGTTCCGGCTGAGATGATAGCCAGGCGGTGATGACAC	4877
Qy	2881	CTCGGTCTTGGGTCACGAGCGCGCGCCCTGTCTGGGCGAGGTTCGAGATGCGCGTC	2940
Db	4878	CTCGGTCTTGGGTCACGAGCGCGCGCCCTGTCTGGGCGAGGTTCGAGATGCGCGTC	4937
Qy	2941	GACGCGGCGGGCTCAGGTTCAATGATGTCACTGCGCGCTGGGATGTGTCCGACGAC	3000
Db	4938	GACGCGGCGGGCTCAGGTTCAATGATGTCACTGCGCGCTGGGATGTGTCCGACGAC	4997
Qy	3001	CTGCGGGGAAAGCCCAACCTTCGCTGTGTCCGAGCGAGTGGCGGGCGGATGCTC	3060

Dh 4998 CTGCGGAGAAACCCAACTCTGCTGCTCGAGGCGAGTGGCGCGGCGCATGTC 5057
Qy 3061 GCCGTGGGCGAGGCGGTGAACGAGCTGTGTGGAGCAACCGATCATGCGCTTTCGAGC 3120
Dh 5058 GCGGTGGGCGAGGCGGTGAACGAGCTGTGTGGAGCAACCGATCATGCGCTTTCGAGC 5117
Qy 3121 GAGAGCTTGTGATCCAGTCAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
Dh 5118 GAGAGCTTGTGATCCAGTCAACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5177
Qy 3181 CTCTCGGCGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3240
Dh 5178 CTCTCGGCGATCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5237
Qy 3241 GACAGATTAAGCCCGCTTTCAGCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3300
Dh 5238 GACAGATTAAGCCCGCTTTCAGCGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5297
Qy 3301 GTCGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3360
Dh 5298 GTCGCTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5357
Qy 3361 GGCACGCGCGGAGAAACGCGCTTTCAGCGCGGAGGCGGCGGCGGCGGCGGCGGCGG 3420
Dh 5358 GGCACGCGCGGAGAAACGCGCTTTCAGCGCGGAGGCGGCGGCGGCGGCGGCGGCGG 5417
Qy 3421 CGCTCGGAGCGGCTGT 3480
Dh 5418 CGCTCGGAGCGGCTGT 5477
Qy 3481 GTGCTCAACTCGCTTCGCGGCGAGCTGATCGAACAAGATTCAATCTCTGCGATCGAC 3540
Dh 5478 GTGCTCAACTCGCTTCGCGGCGAGCTGATCGAACAAGATTCAATCTCTGCGATCGAC 5537
Qy 3541 GGCCTGTTTGTGAGCTTCGCGGAGCGGAGCTTTCAGCGGAGTACAGCTTCGCGTCCG 3600
Dh 5538 GGCCTGTTTGTGAGCTTCGCGGAGCGGAGCTTTCAGCGGAGTACAGCTTCGCGTCCG 5597
Qy 3601 CCGTTCCTGCGCAATCT 3660
Dh 5598 CCGTTCCTGCGCAATCT 5657
Qy 3661 CCGGCGGCGGCTCGT 3720
Dh 5658 CCGGCGGCGGCTCGT 5717
Qy 3721 ACCCTTCCCCCATCGGAGGCTCCGATCGCTCGCTGTGTGTGTGTGTGTGTGTGTGT 3780
Dh 5718 ACCCTTCCCCCATCGGAGGCTCCGATCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 5777
Qy 3781 GCGCAGGCGGAGCATCTTGTGGAGAGCTGTACTCAAGCTGGGTGACCGGAGGTCCAGATC 3840
Dh 5778 GCGCAGGCGGAGCATCTTGTGGAGAGCTGTACTCAAGCTGGGTGACCGGAGGTCCAGATC 5837
Qy 3841 CGATTTCGAACCCAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3900
Dh 5838 CGATTTCGAACCCAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5897
Qy 3901 CTGCGCTGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3960
Dh 5898 CTGCGCTGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5957
Qy 3961 GTCCTCGAGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4020
Dh 5958 GTCCTCGAGGTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6017
Qy 4021 CTGCGCATGATGCTGCTATGAGCGGTGAGAGCTGCGCATCTGATCGAGCGAGCGCTCAG 4080
Dh 6018 CTGCGCATGATGCTGCTATGAGCGGTGAGAGCTGCGCATCTGATCGAGCGAGCGCTCAG 6077
Qy 4081 CTGAGAGCTGTGAGAGAGTTCCTGTCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGG 4140
Dh 6078 CTGAGAGCTGTGAGAGAGTTCCTGTCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGG 6137

Qy 4141 CTGTTGATGCTCTCGCAGAGCTCTCTCTTGTGAGCGGTGTGCGGCGGAGAACTTACGG 4200
Dh 6138 CTGTTGATGCTCTCGCAGAGCTCTCTCTTGTGAGCGGTGTGCGGCGGAGAACTTACGG 6197
Qy 4201 GCAGGCGTGAACAGACTTGTCTCATCGGCGGCGGAGATCAAGACTTGAATATTGCGC 4260
Dh 6198 GCAGGCGTGAACAGACTTGTCTCATCGGCGGCGGAGATCAAGACTTGAATATTGCGC 6257
Qy 4261 CTATGA 4266
Dh 6258 CTATGA 6263

RESULT 9
AR172664
LOCUS AR172664 71989 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 2 from patent US 6303342.
ACCESSION AR172664
VERSION AR172664.1 GI:17912155
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 71989)
AUTHORS Julien,B., Katz,L., Khosla,C. and Tang,L.
TITLE Recombinant methods and materials for producing epothenones C and D
JOURNAL Patent: US 6303342-A 2 16-Oct-2001;
FEATURES
source 1. 71989
BASE COUNT 10108 a 23531 c 26617 g 11731 t 2 others
ORIGIN

Query Match 98.2%; Score 4187.6; DB 6; Length 71989;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 4217; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Qy 1 GTGCGGATGCTGCCATCGAGCGGCGGAGAGATCCGATGCGATGCTCGAGCGAGT 60
Dh 1998 GTGCGGATGCTGCCATCGAGCGGCGGAGAGATCCGATGCGATGCTCGAGCGGCG 2057
Qy 61 TGCCTGTGCGCGGAGTCCCGCGGAGAGCGCTGGATGAGCGCTTCTGAGCGCTCTCGAGGCTCG 120
Dh 2058 TGCCTGTGCGCGGAGTCCCGCGGAGAGCGCTGGATGAGCGCTTCTGAGCGCTCTCGAGGCTCG 2117
Qy 121 CGGACACCGTGGGCGAGTCCCGCGGAGAGCGCTGGATGAGCGCTTCTGAGCGCTCTCGAGGCTCG 180
Dh 2118 CGGACACCGTGGGCGAGTCCCGCGGAGAGCGCTGGATGAGCGCTTCTGAGCGCTCTCGAGGCTCG 2177
Qy 181 GACCCGATGCGCGGCGGAGAGCGCGGATGAGCGCGGATCTTCTGAGCGAGTACGCC 240
Dh 2178 GACCCGATGCGCGGCGGAGAGCGCGGATGAGCGCGGATCTTCTGAGCGAGTACGCC 2237
Qy 241 TGCCTGACCGCT 300
Dh 2238 TGCCTGACCGCT 2297
Qy 301 CATGACTCTTGTGAGAGTGTCTGAGAGCGCTTGAAGAACCGCGGATGCTCTCATTCG 360
Dh 2298 CATGACTCTTGTGAGAGTGTCTGAGAGCGCTTGAAGAACCGCGGATGCTCTCATTCG 2357
Qy 361 GCGCTGTGCTGAGTCCGAGAAAGGAGTGTCTATGAGGATCGGCGGCGGCGGCGGCGGCGG 420
Dh 2358 GCGCTGTGCTGAGTCCGAGAAAGGAGTGTCTATGAGGATCGGCGGCGGCGGCGGCGGCGG 2417
Qy 421 GCGCTGTGCGGAGAGCGGCGTCCGAGAGATGAGCGCTTCAAGCGGCGGCGGCGGCGGCGG 480
Dh 2418 GCGCTGTGCGGAGAGCGGCGTCCGAGAGATGAGCGCTTCAAGCGGCGGCGGCGGCGGCGG 2477
Qy 481 CCGAGGCTGAGAGCGGCGGAGATCTGTATGCTTGTGAGGCGGCGGCGGCGGCGGCGGCGG 540
Dh 2478 CCGAGGCTGAGAGCGGCGGAGATCTGTATGCTTGTGAGGCGGCGGCGGCGGCGGCGGCGG 2537

Oy	541	GTGATATAGCGCTATTTGCTCTTCCGTGGTGGCCGTTCACTGGGCTGTGACAGACTTTCGC	600
Db	2538	GTGATATAGCGCTATTTGCTCTCTCTGCTGTGGCCGTTCACTGGGCTGTGACAGACTTTCGC	2597
Oy	601	TCCGGAGCAATCTCCACGGACCCTGGCTGTGGGATATGCTGATGTTGTCCGACGACC	660
Db	2598	TCCGGAGCAATCTCCACGGACCCTGGCTGTGGGATATGCTGATGTTGTCCGACGACC	2657
Oy	661	CTCGTGTGGCTCTCGAAGACCCCGGCGCTGGCCACAGGACGGTCCCTGCGAAGCAATTTTCG	720
Db	2658	CTCGTGTGTGGCTCTCGAAGACCCCGCGCTGGCCACAGGACGGTCCCTGCGAAGCAATTTTCG	2717
Oy	721	GCGGAGCGCGATATGGGTTTCCGGAACGAGGCGCAAGGGTACCCCGTCTGTGTCTTCAACCGCTC	780
Db	2718	GCGGAGCGCGATATGGGTTTCCGGAACGAGGCGCAAGGGTACCCCGTCTGTGTCTTCAACCGCGCTC	2777
Oy	781	AGTGGAGCCCGCCGCGACCGCGGATTCGATATTTGGCGGGATTCGAGATTCGCGCATCAAT	840
Db	2778	AGTGGAGCCCGCCGCGACCGCGGACCGCGGATATTTGGCGGGATTCGAGATTCGCGCATCAAT	2837
Oy	841	CACGACGGTGGCAGCAGCGGATCTGACCGGTGCCAAACGGGACCTCCCAAGAAATGTTCTG	900
Db	2838	CACGACGGAGCGAGCAGCGGATCTGACCGGTGCCAAACGGGACCTCCCAAGAAATGTTCTG	2897
Oy	901	AAACGGGCGCTTGGCGGAGCAGGAGCTGTGCCCGCGCTTCTGGTGGGTTATGTCCAGGACAC	960
Db	2898	AAACGGGCGCTTGGCGGAGCAGGAGCTGTGCCCGCGCTTCTGGTGGGTTATGTCCAGGACAC	2957
Oy	961	GGCAACGGGACAGACGCTTGTGTGACCCCATTCGAAATTCGAAGCTCTGAATGCGGTATACGGC	1020
Db	2958	GGCAACGGGACAGACGCTTGTGTGACCCCATTCGAAATTCGAAGCTCTGAATGCGGTATACGGC	3017
Oy	1021	CTCGGGCGAGATGTGCGCACCGCCGCTGCTGATCGGGTCTGTGAAGAACCAACCTTTGGCCAT	1080
Db	3018	CTCGGGCGAGACGTGCGCACCGCCGCTGCTGATCGGGTCTGTGAAGAACCAACCTTTGGCCAT	3077
Oy	1081	CCTGAGTATGTGCTCGGGGATCATCTGAGGCTGCTGAAGGTCGTTCTTCCCTTACACAGGG	1140
Db	3078	CCTGAGTATGTGCTCGGGGATCATCTGAGGCTGCTGAAGGTCGTTCTTCCCTTACACAGGG	3137
Oy	1141	CAGATTCTCTGGCGACCTCCACGCGCAGCGCGCTGAACCCCGCGGATCTCATGSGGATGATCTT	1200
Db	3138	CAGATTCTCTGGCGACCTCCACGCGCAGCGCGCTGAACCCCGCGGATCTCATGSGGATGATCTT	3197
Oy	1201	CGGCTGACCGTCAACGCGCGCCCGGACACCGTGTGCCGACTTGAATACGCGCGACCGGCG	1260
Db	3198	CGGCTGACCGTCAACGCGCGCCCGGACACCGTGTGCCGACTTGAATACGCGCGACCGGCG	3257
Oy	1261	GGGGTGAAGCTGTTTGGGATATGAGCGGGACCAAACGCGACGTTGTGCTGTGAAGAGGCGCG	1320
Db	3258	GGGGTGAAGCTGTTTGGGATATGAGCGGGACCAAACGCGACGTTGTGCTGTGAAGAGGCGCG	3317
Oy	1321	GCGGCGAGCTGACACCGCCGCGCGCCGAGCGACCGGCGAGAGCTGTGTCTGTCCGCA	1380
Db	3318	GCGGCGAGCTGACACCGCCGCGCGCCGAGCGGAGCGGCGAGAGCTGTGTCTGTCCGCA	3377
Oy	1381	AGGACCGGCGTCAAGCCCTTGAATGCAACGAGCGGCGCGGCTGCGGACCAATCTGAGAACCTTAC	1440
Db	3378	AGGACCGGCGGAGCCTTGAATGCAACCGGCGCGGCTGCGGACCAATCTGAGAACCTTAC	3437
Oy	1441	CCTTGGCAGTGTCTGGGAGGATGTGTGCGTTCACTCTGTGGAGACGACGCGCGCATGTGAG	1500
Db	3438	CCTTGGCAGTGTCTGGGAGGATGTGTGCGTTCACTCTGTGGAGACGACGCGCGCATGTGAG	3497
Oy	1501	CACCGGCTCGCGGTTGGCGGACGTCGAGGGAGGGGCTGCGGGACGCTTGAACGCTTCG	1560
Db	3498	CACCGGCTCGCGGTTGGCGGACGTCGAGGGAGGGGCTGCGGGACGCTTGAACGCTTCG	3557
Oy	1561	GCGCAGGGACAGACGTCGCGCGGTCGGTGTGCGCAGATATCGCCGATTCCTCAACGCGCAAG	1620
Db	3558	GCGCAGGGACAGACGTCGCGCGGTCGGTGTGCGCGGATATCGCCGATTCCTCAACGCGCAAG	3617

QY	1621	CTCGCTTCTCTTCAACCGGACAGAGGGGGCCACACCTGTGGGCAATGGGCGGTGGCTGTAC	1680
Db	3618	CTCGCTTCTCTTCAACCGGACAGAGGGGGCCACACCTGTGGGCAATGGGCGGTGGCTGTAT	3677
QY	1661	GATGATATGATCGGGGTTTCCGAGAGGCGTTGACCTGTGGCTGAGAGCGTTTCAACAGAG	1740
Db	3678	GATGATATGATCGGGGTTTCCGAGAGGCGTTGACCTGTGGCTGAGAGCGTTTCAACAGAG	3737
QY	1741	CTCGACCGGGCGCTCCGCGAGGTGATGTGGGCCGAAACGGCCAGACGTCGACGCGCGCTG	1800
Db	3728	CTCGACCGGGCGCTCCGCGAGGTGATGTGGGCCGAAACGGCCAGACGTCGACGCGCGCTG	3797
QY	1801	CTCGAACCAAGACGCTTCAACCCAGCGCGCGCTGTTCACCTTCCGAATATGGCTGTCCCGG	1860
Db	3798	CTCGAACCAAGACGCTTCAACCCAGCGCGCGCTGTTCACCTTCCGAATATGGCTGTCCCGG	3857
QY	1861	CTGTGACGGATGTATGGGGGTATGAGACCGGAGTTGGTTCGCGGACATATACATCGGTGACTG	1920
Db	3858	CTGTGACGGATGTATGGGGGTATGAGACCGGAGTTGGTTCGCTGACATATACATCGGTGACTG	3917
QY	1921	GTGGCTGCTGCTCGTGGCGGGCGGTGTTCCTCGCTTGAGACCGAGTGTCTGTGTGTGCTCG	1980
Db	3918	GTGGCTGCTGCTCGTGGCGGGCGGTGTTCCTCGCTTGAGACCGAGTGTTCCTGTGTGTGCTCG	3977
QY	1991	CGCGGGCGCGCTGATGACAGGCGCTCTCGCGCGCGCGCGCGCATATGTTGTGATACAGAGCGCG	2040
Db	3978	CGCGGGCGCGCTGATGACAGGCGCTCTCGCGCGCGCGCGCGCATATGTTGTGATACAGAGCGCG	4037
QY	2041	GAGGCGCATATGAGCTGTGGCTGCGGTGGCGCGCGCGCGACAGCGTCCGGTGTGATTCGCGCGGCT	2100
Db	4038	GAGGCGCATATGAGCTGTGGCTGCGGTGGCGCGCGCGCGACAGCGTCCGGTGTGATTCGCGCGGCT	4097
QY	2101	AACGCTCCGAGCAAGATGTGATCGCGGGCGCCGGGCAACCCGTGATATGCGCGCG	2160
Db	4098	AACGCTCCGAGCAAGATGTGATCGCGGGCGCCGGGCAACCCGTGATATGCGCGCGCG	4157
QY	2161	GCGATGCGCGCGCGCGGGGGCGGAAACCAAGCGCTTCAAGCTTGTGCAATGCGTTTCACTCA	2220
Db	4158	GCGATGCGCGCGCGCGGGGGCGGAAACCAAGCGCTTCAAGCTTGTGCAATGCGTTTCACTCA	4217
QY	2221	CGGCTCAATAGGCGCGGATGTGGAAGCGGTTGGGGGTGTGGCCGAGTGGGTGAGCTACCGG	2280
Db	4218	CGGCTCAATAGGCGCGGATGTGGAAGCGGTTGGGGGTGTGGCCGAGTGGGTGAGCTACCGG	4277
QY	2281	CGGCGGTGATGCTGCTGTGTGACGAATCTGAGCGGAGAGCTTTCACAGACGAGGTGAGC	2340
Db	4278	CGGCGGTGATGCTGCTGTGTGACGAATCTGAGCGGAGAGCTTTCACAGACGAGGTGAGC	4337
QY	2341	TGCGCGGAGCTATTTGGGTATGCGCACGCGCGAGAGGTGTGTGCTTCCGGAATGGAATGTAAG	2400
Db	4338	TGCGCGGAGCTATTTGGGTATGCGCACGCGCGAGAGGTGTGTGCTTCCGGAATGGAATGTAAG	4397
QY	2401	GCGCTGACAGCGGGCGGATGCGGGGACCTTGTGTGAGGTGCGGTCCGAATATGAGCGCTGCTC	2460
Db	4398	GCGCTGACAGCGGGCGGATGCGGGGACCTTGTGTGAGGTGCGGTCCGAATATGAGCGCTGCTC	4457
QY	2461	GCGCTGAGGCTGTGCTGTGATCCCGAGACCGCGCGCGCTGTGCTCGCATGTCGCGCGCT	2520
Db	4458	GCGCTGAGGCTGTGCTGTGATCCCGAGACCGCGCGCGCGCTGTGCTCGCATGTCGCGCGCT	4517
QY	2521	GGGCGTGAAGAGCGGCGGACCGTGTCTGAGGCGCTCGCGGGGCTGTGGGCGGTGTGGTGTGC	2580
Db	4518	GGGCGTGAAGAGCGGCGGACCGTGTCTGAGGCGCTCGCGGGGCTGTGGGCGGTGTGGTGTGC	4577
QY	2581	CTGTGTCTGTATGGGCGCGGCTCTTCCCTCAAGGGGGGGGGGGGGTGGCGGCTGCCAGCTAC	2640
Db	4578	CTGTGTCTGTATGGGCGCGGCTCTTCCCTCAAGGGGGGGGGGGGGTGGCGGCTGCCAGCTAC	4637
QY	2641	CTTTTGGACGCGGACGCTACTGTGATGACACGAAACCGACGACGCGCGCGCGTGTGGAC	2700
Db	4638	CTTTTGGACGCGGACGCTACTGTGATGACACGAAACCGACGACGCGCGCGCGTGTGGAC	4697
QY	2701	CGCCGCTGCTCCGAGAGCGGGTTCAGACGAGGTTCAGAGAGGGGGGCGCGGTGCCGCGCGGC	2760

Db	4698	CGCGGTGCTCCGGGAGCGGGTTCACGACGAGTCCAGAAAGGGGGGCGCGGTGCGCGCGGC	4757
Qy	2761	GACCGGGCGAGCGCTCGGCTTCGACCATTCGCGCGCCGAGCGAGCGCGCGGAGAAAGTTC	2820
Db	4758	GACCGGCGACGCGCTCGGCTCGACCATTCGCGCGCCCGAGCGAGCGCGCGGAGAAAGTTC	481.7
Qy	2821	GAGGCGCGCGGCGACCGTTCGTTCCGTTCGGCTTCGAGATTCATGAGCCAGGCGTGTTCGATAC	2880
Db	4818	GAGGCGCGCGGCGACCGTTCGTTCCGTTCGAGATTCGATGAGCCAGGCGTGTTCGATGCG	4877
Qy	2881	CTCGGTCTCGGGTTCACGAGCGGGCGGCGCCCTGGTTCGGGCGAGGTTCGAAATGCGCGTC	294.0
Db	4878	CTGGTGTCTTCGAGTTCACGAGCGGGCGGCGCCCTGGTTCGAGAGGTTCGAAATGCGCGTC	4937
Qy	2941	GACGCGGCGGGGCTTCAGCTTCATGATGTTCAGACTCGGCTCGGCGATGTCGCCGACAC	3000
Db	4938	GACGCGGCGGGGCTTCAGCTTCATGATGTTCAGAGCTCGGCGTGGGATGTGTCCGACGAC	4997
Qy	3001	CTGCGCGGAAAGCCCAACCTTCGCTGCTGCTCGAGGCGAGTGGCGCCGCGCATGCTC	3066
Db	4998	CTGCGCGGAAAGCCCAACCTTCGCTGCTGCTCGAGGCGAGTGGCGCCGCGCATGCTC	5057
Qy	3061	GCCGTGGGCGAGGGCGTAAAGGCGCTGTGGTGGGCGAACCGGTCATGCGCTTTCGGCG	3120
Db	5058	GCCGTGGGCGAGGGCGTAAAGGCGCTGTGGTGGGCGAACCGGTCATGCGCTTTCGGCG	5117
Qy	3121	GGAGCGTTTGTCTACCAACGTACACACGTCGGCTGGTGTCTGCTCGGCTTCAGGCG	3180
Db	5118	GGAGCGTTTGTCTACCAACGTACACACGTCGGCTGGTGTCTGCTCGGCTTCAGGCG	5177
Qy	3181	CTCTCGCGGATTCGAGCGGCGCCGACATGCGCTCGGTACTTACCGGATGATGCGCGTTC	3240
Db	5178	CTCTCGCGGATTCGAGCGGCGCCGACATGCGCTCGGTACTTACCGGATGATGCGCGTTC	5237
Qy	3241	GACAGAAATGCGCGCGCTTCAGCGCGGGGAGGGGGTGCATTCGATCGGCGGACCGGCGGG	3300
Db	5238	GACAGAAATGCGCGCGCTTCAGCGCGGGGAGGGGGTGCATTCGATCGGCGGACCGGCGGG	5297
Qy	3301	GTCGGTCTCGCGCGGTCGACGTGGGCGGAGACGTCGGGAGCGGAGTTCATTCGACGCGC	3360
Db	5298	GTCGGTCTCGCGCGGTCGACGTGGGCGGAGACGTCGGGAGCGGAGTTCATTCGACGCGC	5357
Qy	3361	GGCACGCGCCGAGAAACGCGCCTTACCTTGAGGTGCTGGGCGTGGCGGTATGTGAGCGATTC	3420
Db	5358	GGCACGCGCCGAGAAACGCGCCTTACCTTGAGGTGCTGGGCGTGGCGGTATGTGAGCGATTC	5417
Qy	3421	CGCTTCGACCGGTTTCGTTCGCGGACGTGGCGCGCGTGTGACGCGGCGCGAGGAGTACGTC	3480
Db	5418	CGCTTCGACCGGTTTCGTTCGCGGACGTGGCGCGCGTGTGACGCGGCGCGAGGAGTACGTC	5477
Qy	3481	GTCGTCAACTGCTCTCGGCGGAGCTGATTCGACAAAGATTTCAATCTCTGTCGATTCGAC	3540
Db	5478	GTCGTCAACTGCTCTCGGCGGAGCTGATTCGACAAAGATTTCAATCTCTGTCGATTCGAC	5537
Qy	3541	GGCGCGTTTGTGAGCTTCGGCAAGCGCGACTGTTTACCGCGATTAACGACTCGGCGTGGG	3600
Db	5538	GGCGCGTTTGTGAGCTTCGGCAAGCGCGACTGTTTACCGCGATTAACGACTCGGCGTGGG	5597
Qy	3601	CCGTTTCTCGGCGCATTTCTCTCTTCTCGTGGTGGATCTTCGCGGGGATGATGCTTGAAGCG	3660
Db	5598	CCGTTTCTCGGCGCATTTCTCTCTTCTCGTGGTGGATCTTCGCGGGGATGATGCTTGAAGCG	5657
Qy	3661	CCGCGCGCGGTCGCTGCGCTCTTGGAGAGGCTCCCGGCTCGATTCGCGGCGAGGCGTGTTC	3720
Db	5658	CCGCGCGCGGTCGCTGCGCTCTTGGAGAGGCTCCCGGCTCGATTCGCGGCGAGGCGTGTTC	5717
Qy	3721	ACCCCTTCCTCCCATTCGCGACGCTCCCGATTCGCGCCGTGTTCGCGCATTCGCTTCGAGCATG	3780
Db	5718	ACCCCTTCCTCCCATTCGCGACGCTCCCGATTCGCGCCGTGTTCGCGCATTCGCTTCGAGCATG	5777
Qy	3781	GCGGAGGCGGACATTTGGGAGCTGTAATCACGCTGGGTGACCGGAGGTCGACATTC	3840

[illegible]

Db 254 GCTCGTGGCGCATCCACTTTCCTGCGCGAGCGCTCGCGCGATCTCGCGCT 313
Qy 624 GCGCTGCGGATATCGCTGATGTTGTCGCCAGCACCCCTGTGTGCTTCGAAAGACCG 683
Db 314 GCGCGGAGGCGGCAATGCTCTTTCGTCGAAGAGATGATGATGCTGGGCGCATCCA 373
Qy 684 GCGCGTGGCGAGGAGCGTTCGCTGCAAGGCAATTTTCGCGGAGCGCGATGCGTTTGGAG 743
Db 374 GCGCGTGTGCGCGGATGCGCACTGCGCGACATTCAGCGCTCGCGCAAGGGTTGTGTGG 433
Qy 744 AGCGCAAGAGATGCGCGCTGCTGCTCTCAAGCGGCTCAGTGGAGCCCGCGCGAGCGGCA 803
Db 434 TGGGAGAGGCGCTCGGATGATGCTGTCTCAAGCGCTCTCCGACGCCAGCGACGGGCA 493
Qy 804 TCGGATATTGCGCGTGTATCGAGATTCGCGATCAATCAAGCGGTGCGAGACGGGTCT 863
Db 494 TCGGATCTGCGCTGATTCGGGGTTCGGCATGATGAGATGCGCGGTTCGACAGGGTT 553
Qy 864 GACCGTGGCGGAGCGGAGCTCCGAGAGAAATCGTGTGAGAAACGGGCGCTGGCGAGCGAG 923
Db 554 GATGACACCCAAATGTGCTGCTCAGAGAGCGCTTTCGCGAGCGCTGCGAGAGCGCTCG 613
Qy 924 CTGCGCGCGCTTTCGCTGAGTTATGTGAGGCAACGCGACGCGGACGACGCTTGTGTA 983
Db 614 CGTCAACGCGGGGCGCATGCGTTATGTGAGACCAAGGAGACCGGGACCTCGCTGGGCA 673
Qy 984 CCGCATTCGAATTCAGGCTTGTGATGCGGTATACGGGCTTCGGGCGAGATGTGCGACGCG 1043
Db 674 CCGGATTCGAGGTTCGAGGCGCTGCTGTCGTTGGGGCGGCGCGGCGAGTGGAGCGG 733
Qy 1044 GCTGCGATTCGGGTGCGTGAAGAACCAACCTTGGCCATCTGATGATGCTGCGGAGTCC 1103
Db 734 CTGCGTGTGCGGCGAGTGAAGAACCAACTTCGCGCACTGAGAGGCGCTGCGAGGCGTGC 793
Qy 1104 TGGGCTGTGAAGTGTGCTTTCCTTCAGACGCGGCGAGATTCTTCGCACTTCACGCG 1163
Db 794 GGGTTTGAATCAAGGGGCGCGCTGCTGTGACCAAGATGATTCGCGGAAACCTTCATT 853
Qy 1164 GAGGCGCTGAACCCCGGATTCATGAGGTTGATCTTCGCTGACCGTCAAGCGCGCGCG 1223
Db 854 CCACACGCTCAATCCGCGGATCCGGATCGAGGGGACCGCGCTCGCGTGGCGACGAGACG 913
Qy 1224 GACACGTTGGCGGAGCTGGAATACCGCGCGAGCGCGGGGGGAGCTGCTTGGGAGTGA 1283
Db 914 GGTGCTGTGCGCGCGGCGGCGGACCGCGCTTCGCGGGGTTGAGCGGTTGCGCTCAG 973
Qy 1284 CCGGACCAACGCGCAGTGTGCTGGAAGAGCGCGCGGCGAGCGTGCACACCGCGCG 1343
Db 974 CCGGACCAACGTCATGTTGTGCTGGAAGAGCGCGCGGCGAGGTTCTGCAACCGCGAC 1033
Qy 1344 GCGGAGCGACCGGAGAGCTGCTGTGCTGTGCGGACGAGACCGGCTCAGCCTGTGATGC 1403
Db 1034 GCGCGGCGGCTCAGCGGAGCTTTCGCTGTGCGGAGAGCGCGCGCGCTGACGCGC 1093
Qy 1404 ACAGGCGGCGGCGGCTGCGGAGCATCTGAGACCTTACCTTGCACTGTCTGAGGAGAT 1463
Db 1094 ACAGGCGGCGGCGGCTTCTCAGCGCAATCGCGGATACCGGAGCAGAGGTTCTGAGACG 1153
Qy 1464 GCGCTTCACTTTCGCGACGACCGGAGCGGAGTGAAGCACCGGCTTCGCGTGGCGCGAC 1523
Db 1154 CCGCTTCAAGCTGTGATGAGCGCGTGAACCGGATGAGCACCGGCTTCGCGTGGCGGAC 1213
Qy 1524 GTGAGAGGAGGCGCTGCGGAGCGCTGTGAGCGCTGCGCGGCGAGGACAGACGTCGCGG 1583
Db 1214 CTGCGCGGAGGCGCTGCGAGGCGCTGTGAGGTTGCGGCGGAGGACAGACCGCGGAG 1273
Qy 1584 TGGCGTGGCGATATTCGCGGATTCCTCAAGCGGCAAGCTGTGCTTTCCTTCAAGGACA 1643
Db 1274 CCGCGCGCGGCGGAGGCGGCTTCTCGCGCGGCAAGCTGTGCTTTCCTTTCGCGGAG 1333
Qy 1644 GCGGCGCGAGACGCTGAGGATGAGCGCTGTGAGCTGTATGATGATGCTGCGCTTCGCG 1703
Db 1334 GCGGCGCGAGGTCGCGGAGATGGGCGGTGTGTGGAGGCGTGGCGCGGCTTTCGCGCA 1393

Qy 1704 GCGCTTCAACCTGTGCGTGAAGGCTGTTCAACAGAGACTCGACCGGCGCGCTTCGCGAG 1763
Db 1394 GACCTTCAACCGGTCGCTCAAGCTCTTTCGACCGGAGCTTCATAGCGGCTGTGAGAGT 1453
Qy 1764 GATGAGGCGAGACCGGCGAGGTCGACCGCGCTGTGACAGCAGGCTTTCACCA 1823
Db 1454 GATGAGGCGAGACCGGCGAGGTCGACCGCTGTGTCGACAGAGGCTTTCACCA 1513
Qy 1824 GCGCGCGCTGTTCACCTTTCGAATATGCGCTGCGCGCTGTGCGGCTGTGAG 1883
Db 1514 GCGCGCGCTTTCGCTGAGTACGCGCTGCGCGCTTTCGCGCTGTGAGGCGTGA 1573
Qy 1884 GCGGAGTTGCTGCGCGGCTTACATGATGAGTGTGAGTGTGCTGCTGCGGCGG 1943
Db 1574 GCGGAGCTGTGTGCGGTCATGAGCTGTGAGTGTGAGTGTGCGCTGTGAGGCGGT 1633
Qy 1944 GTTCTGCTTGAAGACGCGGCTGTCTGATGAGTGTGCGCGGCGCTGATGACAGCGCT 2003
Db 1634 GTTCTCTCTCAAGACGCGCTGTGCGCTGTGATGAGTGTGCGCGCGGCTGTGAGGCGCT 1693
Qy 2004 GCGCGCGCGCGGCGATGATGATGATGAGGCGCGGAGCGCATGTGCTGTGCGCT 2063
Db 1694 GCGCGCGCGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1753
Qy 2064 GCGCGCGCACGAGGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2123
Db 1754 GCGCGCGCACGAGGCTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1813
Qy 2124 GCGCGCGCGCGGCGCAACCGTGTGATGATGATGATGATGATGATGATGATGATGATG 2183
Db 1814 GCGCGCGCGCGGCGCAACCGTGTGATGATGATGATGATGATGATGATGATGATGATG 1873
Qy 2184 AACCAAGCGCTCAACGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 2243
Db 1874 AACCAAGCGCTCAACGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1933
Qy 2244 GCGCTTTCGCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2303
Db 1934 GCGCTTTCGCGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1993
Qy 2304 CAATGTAGCGGAGGCTTTCAGACAGAGGATGATGATGATGATGATGATGATGATGATG 2363
Db 1994 CAATGTAGCGGAGGCTTTCAGACAGAGGATGATGATGATGATGATGATGATGATGATG 2053
Qy 2364 CCGCGAGAGGTTGTGCTTTCGCGATGATGATGATGATGATGATGATGATGATGATGATG 2423
Db 2054 CCGCGAGAGGTTGTGCTTTCGCGATGATGATGATGATGATGATGATGATGATGATGATG 2113
Qy 2424 CACTTTCGTCAGAGTGTGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2483
Db 2114 CACTTTCGTCAGAGTGTGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2173
Qy 2484 GAGCGCGCGCGCGGCTGTGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2543
Db 2174 GAGTGCAGGCGCGGCTGTGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2233
Qy 2544 GCTGAGGCGCTTCGCGGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2603
Db 2234 GCTGAGGCGCTTCGCGGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2293
Qy 2604 CCGCTCAGAGGCGCGGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2663
Db 2294 CCGCTCAGAGGCGCGGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2353
Qy 2664 GATGACAGAGGCGCGGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2723
Db 2354 GATGACAGGCGCGGCTTTCGAGAAATGACGCTGTGCGGCTGTGCTGTGATGCC 2413
Qy 2724 CGAGAGGTCGAGAGG 2740
Db 2414 CCGCTTTCGAGGTCGAGG 2430

RESULT 11
AX024276 10910 bp DNA Linear PAT 24-NOV-2000
LOCUS Sequence 81 from Patent DE19846493.
DEFINITION AX024276 AX024383
ACCESSION AX024276.1 GI:10184550
VERSION
KEYWORDS
SOURCE
ORGANISM
Polysangium celluloseum
Polysangium celluloseum
Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
Sorangineae; Polysangiaceae; Polysangium.
REFERENCE
1
AUTHORS
Beyer, S. and Mueller, R.J.
JOURNAL
Patent: DE 19846493-A 81 13-APR-2000;
BIOTECHNOLOG FORSCHUNG GMBH (DE)
COMMENT
On Oct 15, 2002 this sequence version replaced gi:10184587.
FEATURES
source
1. 10910
/organism="Polysangium celluloseum"
/mol_type="genomic DNA"
/db_xref="taxon:56"
BASE COUNT 1852 a 3124 c 3613 g 2321 t
ORIGIN
Query Match 21.8%; Score 928.8; DB 6; Length 10910;
Best Local Similarity 60.5%; Pred. No. 3,3e-92;
Matches 1606; Conservative 0; Mismatches 1027; Indels 21; Gaps 4;
QY 25 GCACCCGAAAGATCCGATTCGATGTCGAGCGAGTTGCCGCTGCCGCGTGGATTC 84
DB 2244 GCGCGGATGACGATCGCCATCGTGGTCCCGCTGCCGTTCCAGAGGAGATGAG 2303
QY 85 GATCGAGCGGATTCGAGCGCTCTCGAGGGCTCGCGGACACCGTCGCGGAGTCC 144
DB 2304 GCGCTGAGACATCTGCGCGCATCTGCGCCGAGGACATGTCGACGACCGAGTGC 2363
QY 145 GCCGAAAGCTGGAGTGAAGAGCGTGTGATCCCGACCCGATGCGCCGAGGAGA 204
DB 2364 GCGGACGCGTGGCGCGCGGCGGAGCTGTGACGCCCATCGGAGTTCGCGGCG 2423
QY 205 CCGCTTACGCGCGCATCTTCTCTAGCGACGCTAGCTGCTTGAAGCGCTCTTTC 264
DB 2424 TATGTGCGCAAGGATGCTCTCTCGCGATGTCGACGCTTGAAGTGGCGCTTTC 2483
QY 265 ATCTGCGCTGCGGAGCGCTGCGGATGAGCCCTGACATCGACTCTTCTGCGAGTGC 324
DB 2484 ATTTCCTGCTGAGGAGATGAGCTTGAACCGCAACGCGGCTGTGCTGAGAGTGA 2543
QY 325 TGGAGGCGCTGAGAAAGCGCGCATGCTCATCGCGCTGCTGCGTACGGAACGGA 384
DB 2544 TGGAGGCGATGAGAGCGCGCTGCGGACGACCCGATGCGCTGCGGAGAGCG 2603
QY 385 GTGTTCATCGGAGATCGCGCGCTCGGAATAGAGCGCGCTGCGGACGAGCGCTGC 444
DB 2604 GTGTTCGTCGAGATGAGTGGAGCG--AGCAGCGGAGCGGATGAGAGCGCTG 2660
QY 445 GCAGAGATCGACGCTCATGCGGCGCTGGGAGCATGCCAGCTGCGAGCGGCGGATC 504
DB 2661 GACCGCGCTGCTGTCGTCGAGCACACCGCAACCTGCTCAGCGCTGCGCGAG 2720
QY 505 TCGATATCCCTCGGAGCTGCGAGGCGCGTGTGCTCGCGGATGATAGGCTATTGCT 564
DB 2721 TCGTCTTCTCTGCGCTGCGAGCGCGCCGACGATGACGCTGAGACCGCTGCT 2780
QY 565 CTGGTGGCGCTTCACTGCGCTGTCAGAGCTTGGCTTCGCGGAGATGCTCCAGCG 624
DB 2781 CTGGTGGCGCTTCACTGCGCTGTCAGAGCTTGGCTTGGAGATGAGCAAGCG 2840
QY 625 GTCGTCGCGGATGAGTGTGTCGCGGAGCAACCTGCTGTCGCTCTGCAAGACCG 684
DB 2841 GCGCGCGGCTCAGCGCTGCTTGTGCGCGGATGCTGCTGCTGCGCGCTGCG 2900

QY 685 GCGCTGCGCAGGACGCTGCTGCAAGCATTTTCGCGGAGCGGATGAGTTCGAGCA 744
DB 2901 TTGCTTTTCCCAATGAGCGGCTGCAAGCATTTTCGCGGAGCGGCTTTCGCGG 2960
QY 745 GCGGAAAGGTCGCGCTGCTGCTTCTTCAAGCGCTGCTGAGCGCGCGGAGCGGAT 804
DB 2961 GCGGAGGCTGCGCGCTGCTGCTTCAAGCGCTGCTGAGCGCGCGGAGCGGAG 3020
QY 805 CGGATATTCGCGGATTCGAGATTCGAGATTCGAGATTCGAGATTCGAGATTCG 864
DB 3021 CCATCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3080
QY 865 ACCGTGCGCAACGAGCTTCCAGAAATGCTGTAAGCGGCGCTTGGCGGAGCG 924
DB 3081 ACCGTGCGCAACGAGCTTCCAGAGCGCTTGTGAGCGGCGCTTGGCGGAGCG 3140
QY 925 TGGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 984
DB 3141 GTGCGCGCGCGCGAGGTCGATTTCTGAGAGTGCACGAGGAGCGGAGCGCT 3200
QY 985 CCATCGAAATCCAGCTTGAATGCGGATTCGAGATTCGAGATTCGAGATTCGAG 1044
DB 3201 CCATCGAGGTCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3260
QY 1045 CTGCTATCGGATTCGCTGGAAGACCAACCTTGGCCATCTGATATGCTGCGG 1104
DB 3261 CTGCTATCGGATTCGCTGGAAGACCAACCTTGGCCATCTGATATGCTGCGG 3320
QY 1105 GGGCTGTAAGTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1164
DB 3321 GGGCTGTAAGTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 3380
QY 1165 CAGGCGCTGAACCCCGGATCTCATGAGGATTCGCTGAGCGCTGAGCGCGCG 1224
DB 3381 GAGGAGCTCAACCGGATCTCATGAGGATTCGCTGAGCGCTGAGCGCGCG 3440
QY 1225 ACAACGTCGCGGATTCGGAATGACCGCGGAGCGGAGCGGAGCGGAGCG 1284
DB 3441 GTCCCTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3500
QY 1285 GGGACCAAGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1344
DB 3501 GGGACCAAGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3560
QY 1345 CCGGAGCGGAGCGGAGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1404
DB 3561 CCGGAGCGGAGCGGAGCGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3620
QY 1405 CAGGCGCGGCGCTGCGGACCATCTGGAAGCTTCCGCTGCTGCTGCTGCT 1464
DB 3621 CAGGCGCGGCGCTGCGGACCATCTGGAAGCTTCCGCTGCTGCTGCTGCT 3680
QY 1465 GCGTTCAGTTCGCGGACCGGACCGGATGAGACCGGCTGCGGTCGCGGAG 1524
DB 3681 GCGTTCAGTTCGCGGACCGGACCGGATGAGACCGGCTGCGGTCGCGGAG 3740
QY 1525 TCGAGGAGGAGGCTGCGGAGCGGCTGCGGAGCGGCTGCGGAGCGGAG 1584
DB 3741 TCGAGGAGGAGGCTGCGGAGCGGCTTTCGCGGAGCGGAGCGGAGCGG 3800
QY 1585 GCGGTCGCGGATTCGCGGATTCCTTCAAGCG--GCAAGCTGCGCTTCTT 1641
DB 3801 GCGGTCGCGGATTCGCGGATTCCTTCAAGCG--GCAAGCTGCGCTTCTT 3860
QY 1642 CAGGCGCGGAGCGCTGCGGATTCGCGGATTCGCGGATTCGCGGATTCG 1701
DB 3861 CAGGCGCGGAGCGCTGCGGATTCGCGGATTCGCGGATTCGCGGATTCG 3920
QY 1702 GAGGCGCTTGAACCTGCTGCTGCTGCTGCTTCAACGAGAGCTTCAACG 1761
DB 3921 GCGGCTGAGGAGGCTTCAACGAGAGCTTCAACGAGAGCTTCAACGAG 3980
QY 1762 GTGATGTGCGCGAAGCGGACGCTGCGAGCGCGCTGCTGCGAGCAAGCTT 1821

[illegible]

SOURCE	genomic DNA
ORGANISM	genus: spix gene; spil gene; spiz gene. polyangium cellulosum polyangium cellulosum Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE AUTHORS TITLE	1 Knauber J., Blocker H., Mueller R., Nordiek G. and Beyer S. The spirangliene synthase from Sorangium cellulosum So cego - Module duplications as strategy behind the evolution of polyketide synthases
JOURNAL	unpublished
REFERENCE	2 (bases 1 to 48000)
AUTHORS	Beyer S.
JOURNAL	Direct Submission Submitted (06-AUG-2002) Beyer S., Nbi-Mx, German Research Centre for Biotechnology, Mascheroder Weg 1, D-38124 Braunschweig, GERMANY
FEATURES	location/Qualifiers
source	1..48000
gene	/organism="Polyangium cellulosum" /mol_type="genomic DNA" /db_xref="taxon:56"
CDS	1..7558 /gene="spig" <1..7558 /gene="spigG" /function="biosynthesis of spirangliene" /codon_start=2 /trnas_table=11 /product="polyketide synthase" /protein_id="CAD3448.1" /db_xref="GI:27528334"
translation="IIGSDKRLASALEAAGAHTRRDPCAMIDAIRGGAAEPPAVL ACTRPDPSPGALAAATHGAADALVALQDALEERLASCALVLTETRAVTRPDEDVL DLVHAPWGLVRAAOVENPDRALIDVADDALESLPALPVLVAARAADAEPLA LREGRVLPRIKIGDGDALVPSSAAYRHHVAKKTELESLAVPHEATRPALBEOI RVAVRAAGNFRDALDGMVPGDPLGAEQGLVLENGPTELEAPAGRWGLISA APRPIALADHRWYSRIPAGLSFVEAASIPVPLETLYAYGVLDIRGSEVLYHAAG GVGIVAAQLARHLGAEPFATSPPKKATTLALGLDDDHIASSSGPPEAFRTSGR GAVGVNLDLSARELVADSLRLVALAPDGRFVEMKSDVRDPLAASHPSVAIFYADLA EADPERIOAMIAELVILFERGVILAPVPIITPCDILRAPOAFALAOALVGVKVLITPR ALDPBGSLVTGGTGTGALVARIHVORHCEVLSVSRGPAPADLRRLLEAAG ASVAYACDAGDRLEERLAVI PADRLPVAAGAAITLDDGVLITSLPERLSVMSA KDAAAALDELTRPLDLSARVPAISLGVSGPDSANVYAAANFLDALAQBRSGRP AVSDMGYMERKGTGLAHLDGTDLOQRARGLPLTTDDGLALFDALAHVDALVA PPEPVALRAHAVAPILIRGLRAGARAKKLANNASAPLAQRLAGLAPEKEREAMIDV VREBI GVILGLPSPRHLDGDRPLKEGLSLVAMELRNLALVTRLPTLLEDFYR PTLARLEIALTALVGRGDAAVPAASARAASGASDDALAIYAMACRPGVLTPEDEYR LLEGGDIA SAPENGGMIDLALDGAARCGGLHADLIDAEFPGISPERTALDPO QRLLEMTWALFRAGIIPASLHSGSKGVTVGSGDYARILHSGSGAQMSELSGQ IGVSGASVASGRITAYALGLEPRAISVDTHCSSLVLAHLACHAIRGECSLAGCV TVAATPGIFGEFAOHAGADGDRCKSFSDANANGSEBGIVLLERLSPARKNGHV LAIVRGAVNOGDKSQGLTLPNGPAQOBYRDLALASGLADPADVDVEAGTCTTLD PIERALFATYGAHAAKERPLMTGSKNSNIGHQAAIGIGVIMVLAOSGLVPTL GMAAAVEITADLEALPEGRALAIATISPPAAVSGPDDIDELVELSGAOLFA RKRVUYASHASHVEALEERTLELDGLIATRPVPSVATGEBRLDGEGLAAHWYR NHRVYRFEHATRALDDGRFFVEVYSPHVLTVALRDTDIDAGHVAVACASIRDEG DLARPNLSTAEHLTRGLAVDMWRAFFAFARFATVLPYARQERFWMLDAKASGVITDA AASPBAERFSAVEDGDAVTDALVDGDERALATLPTLARTRRARARAGSTIDA WRVRWRALATAPKGIAGTILVVAPEVGLADYARALREARGAVLVYPADABDA ARFPLIGRLAVAGGQIRGIVISFALDEBPALDEPSVPLGLAVSGVLOALDADID APIMLYTRGAVVAGSERVTRPQAMTWKIGRVLALBHRPMKGVLDIGSEASPDNG ARDQAHDRILANIVARDEDDOLALRAAGTYRRVLYRAPIHDGKRPAPFARGAALVYG GTQALGHVRLMARI GAEBI VLASRRGAAPALAAELSLAGVATLAACDVADR EAAVALRVDEAGATVRAVFAHGAAMHEPVAMVEELADLIAAKARQAHLDVF AQBPLDFVLFSSGAGWGQOAAVANAAPFDALAEARADGLATSTAMQAMGQ GMALTAERLTKRIGVAPMDPELAVATLAALHAETCLAVVDNARFAPSASAP RPLIDELAEKRSALDAREPDDARVALAAGEPALTPTTLAALDEGRHGLIATVPT ETAAVIGHAASRVEPRRGFFDGLGSLMSVEIKRRVQARTGKILPTTLAHDHPTPSA	

gene
CDS

```

LASKVLAIVLHDTTPRASPAELERLEGMLSAIVADEALRDDLTAIRAFLEDKRAVR
TEREDDAAFEKKGASADELIRLIDQGLDRIDVRY"
/gene="22595
7542..22595
/gene="sp1H"
7542..22595
/gene="sp1H"
/function="biosynthesis of spirangiene"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="polyketide synthase"
/protein_id="Cad3449.1"
/db_xref="GI:27528335"
/translation="MSTVTNDLTLEYRLRLTOELHRSSTRLATEERHREPIAVIGLG
LRLGGLHSDRLTLTFLPEGRDALPIPSHWMNADATDPDPAVGSXYNDAMLR
VLDLADDPFGISPREAKVDRQHRLLLETSMQLEDAIGYPSLRDSKQVFGVSGS
DYALFQSDRDASEAIVMGMSIFPAKSLATLGLQSPALSDTPACSSVYALHACQ
SLRQECDLAVAGVQWSSPEVVLISRTALASDGSKTFSANADGKGEVVVL
AVERLRARAKGRPIIAVIRGSANHDGSSGITVPPNGAQKYLRAALDARLVPD
VDVECHGTGTSIGDPIEVNLAAYVGEGRPKDPLFLGALKTNIGHLEFASGLAIVA
KMAVSHRATPLATLHTSPLNPLVDMDLPIRVVDARPMTRDDGAPRRAGVATGL
SGTAAHIVEEAPAEPEPTPDAPALPAVPVILSGTKDELRAAARIHAHLGRD
ARLYDIAASLATTTHHPRRAVVAADRDELGLDALRGEAPGSVVASITAGVY
VVFPGSGQWGNARALASVGFROBIAICERLAPHVMSLGAVALRGDGEATLL
GRVDVQVPIFAVAVVLAALMRSIGVTPDAVVGSGEIAAAYVAGLSLEDAKVA
LRARLTKIAGRGMAAVELEGARDETEARLEFGDAIAIAINSRPAVLVAGDTAIDA
LVRLEAAOIFPARKVVDYASHAVEIERELADLAGIERAGAVLPVAGVAGLAL
DGSRLDAHMFNRNLRSTYRPEADRALHDGHRFPVSPVVALDELTSIGD
AAVYSLRREBDEGLERFLISAGELHAGHAIIDKAPFLCARVTLPTIAFORERN
LDAGDADDEDAFGASAEETAERAIERGVAAISDLAHVADSGRSALSLAMPALS
AMRSRREOSTVGMRYRVEWRPVSAASRGVAGTWLVAVAGVASDASLAMPALS
RGADVPLALEASDITAHDAASRLDAIAGRPARGVLTPLDTPADLGDPAVAGL
ARSLAIOALAEILAEPIWITRGVAIGRSRIESPVOAMGLKRVSLDEPBA
GGVLDLPRADHGDALDATALERLVAGVSGRGEDELALPGLFARRELAIPGDA
OREPREVTAVTGTGALGANVAMLRHGAHIVLASRGAAPGABALAEITELG
ARVIAACDVADRAIIELFALPAEPEPLIAIFTAGVTDGVLSLTPLELTVLHG
KADARTLHELTKTRDLAEVLFSSAGVFNMAQASVAAANALDLAEERLALGIR
ATVAMGAMAGAMLTEDVAGLRGRGAPEPIAAVAGLSALDRDSTHTVITASIM
ARFASVASARORLVADLPARALAKOPASTEHDELTALVATRLSGEARLHV
SLVIAOTRAVIGHDRSRIDRQGFPLGDSIMAVELRLOKATGVKLPATTTPR
PSPRKLGLDLALPIGGAAPSTARPAPARADASDEITAVIGLIPGSDVDLAL
FRLIEGRDAVAPIPASRDVGAVIDPEPEKGTYYVGHAMLDRLVDFDGFGLIS
REARHVPORHLLETAMQLEAATVPSLRVAGTFVAGASDVAVLOSSEDAE
AYAMGPASFPAGLAETLGLQSPALSDTPACSSVYALHACQSLRQECDLAA
GVQMASPEVVLISRTALASDGSKTFSANADGKGEVVVLALERLDRANRH
PVIAVVGTAVMHDSAGSITAPNGSSQKTRALADARLAPDVVVECHGTGTSI
GDPLEVOALAVIGEERAPERLRLGAVKTNIGHLEFASGLAGVVKIVASIRHGTPLA
TIHTNPNRPHIDMDALPIRVVDAEPERSADGSPRAGVAFGLSGTNHVIYEEAP
EPDAPKARALPLSGALPEVLSAKSDAALAAQAAALDQIARFPDADLVNAAALAT
TRSQDFRAAIIAASDHGALVALLGLAGVAAPOGTFAKKGADLAFLEFGQGORAA
MGRGLDAPFVPRDALDAVAVSHLRELDRLPLRDVLPAPKGSLESLIDREFEPOALF
ALRYALFRIVEAKGVTPVILGHSVGELEAAHVAAGVSLDACTIVAAKRLMALRA
RDGMVTVHATEAVLALLEGDRAEILAIINASTVIVADVAIVARVAFKAR
KATLRVSHAESHMDPMLDAFRRVAEGLTFHPRIPIVSNVTKARLADBEIRPDR
VRHRSVAVRFADGINTLEADGVSSFELGEPGVLSALGESLAANAQVAFVPLRQ
RADVDALTAALSLHVRGHVDMAFPEGPDPOTVAFYFQDRFLDPAFCATRD
EAAAPABEDAEMRAVADGVALGATILTNASGEDHLSALATLALSAMRABEREL
VDARVYVWKEPLTPAARDVAGTWTLLTAGADHARLEIARLITARGAEVITMPLA
PAEDRVRLARVVGALADGELRGIVSFALDITETTLATHTALPAGLALTRLVAGS
DMGLEAPLMTVTRGASVSGRDLASAQOSMTGILGRVGLGHEBERNGVLDVAGDT
IDARLRLVTLTLAGSDGEQLARPTGIFRRLVRAPLEGHPARVITVAGTSVVG
GTGALGAVARFLAGRGHEHLVLSRGEEDAGALAELEAGARTVLAACVAVR
ASVALLARLADAGDPVRAVYHAGVVAQALASTDIDGVAVVAARVAGLHIELL
GDRLEDAFVLPAAGAGVWGGCGAAYAAAGALFLDALVEPRAALGTATSTAMGWAG
GMLDHDADDAVEELRKGVIPNAPALAIALAGLDHGETAVTASDARLFAVRA
SARPRLLDLPEAKRAIEASAPAGAGASADALVEALPSEADARLVELVLA
ETAIVLHADASADIPKGFPLDGLDSVAVELRKIRARVGLPVTIVFDHSGPHR
LATVVRALPLADAPADATTARSSAAHIDEPITAVGLGILPGVEBERGELMF
LEQDRGVGPIPKDRWDADATVDDPEAKGSYVBAALDRVLLPAAFFGISPRBA
KONDORHILLETAMQLEAATVPSLRVAGTFVAGASDVAVLOSSEDAE
AMGTLSFPAKRLAFITGLQSPALSDTPACSSVYALHACQSLRQECDLAAQV
VMTAEPPVLLSRTALADRSKTFSAHAGVGRGCVVVALLERLDAERARVLA
AVNLTAEVNHDSAGSITAPNGTSQKTRALALDAGLAPADVVECHGTGTLGDP

```

```

IEVOALAAVYGEGRBAERPLLGLAVKSNIGHLEASGLAGVAKITVLEHGALEPATIH
SRBNPPIVMDALHVRVVDAAELMARDEBTPRAGVAGLSCGTNNAIIIEPRAR
RGAATDEPPPAVPPVVLASGERLRLAQRARLVHVTGRGDLVDVNAASLATSHP
EHAATVARDLGLLEDAIAGLGRSPAGVIVSGTGGKALLFTGGGGAAMGRGL
YGAFSVFRDAPDAACSTLDBREHRLPDDVVPAPGSLAALIDETITVQTSPLAEVA
LFLVEAMGVTPPHVLGHSIGELVAHAHVAGVLSLEDCITLVAARLRLHRLPDAVA
VAHAASEVSHLDLDNAAGLDGRASIAVNAVPSSTVAGREAVLAAAREALGRKT
SRLRSHAFSHHLDDPMLAEPRVAEIVREPEPIPIVSNVTKARASNDICADYVW
RHARGVRFADGLEALADGVSTLEIEGPHAVLGLAHETLESAPVFLPLRQGRD
VDITLTAIAGLHARGHVMDAPFAPAPNARVALLPYAFORERFLMPPKSSRAA
ADSNGRPLIAGTRLDLPDGSVHTLIDGROQPLADHSTYGRIVYGAHRLAVILAV
AESHWPSPLELHVEFVRALTFEESETAVRVHLLPAGDGEHFAVTAQTQEGM
TTHATATLGAAPSALRHPRAHEPLAGBEPASVAGASALKURLASLATSNS
QTHACGRTOVGLAADGVPTDAPLGLVDNAFVAGSATVGLRPAHNDSTSNV
LPESIRLWVYGALEPCVADFMREIIPDARVIGDITLVRGEASGASGREGT
PIATIEGLARBPADKELELPARNYALTTPQAPASPPRMRMLIGDSIGIAT
GIADADPVDVLDALRGRAPCEVTVACATLTHGSDTDETVSSVTHAARALA
LUNAMLSDEHLASTVLTLQRAVATAGSDVLDLASAALGLVRSQAEERPAIIL
VLDLSDASRRALPYAVAGLDAEPOLARLDGRRLARLRAVAPATANAASTIS
ATPRALPEGTVLTGCTGTGALVARELVRHVKILVLSRGPDPAPGATDLAEL
QARGASVVLAAADAEPALARVLLAIPHDRPLVAVVAGTLDGVLSSWTPARLSA
VRAKVDAVNLDBORHSPRLRAVLFSSLSGVTSPOASVAAANFLDALHARRA
RGUPVAVLDMGYNAQTSGLTKHEQADLRIRARGGLEPLASDEHALFDLADADAC
LVPAFADALRKSGLSALRKLVLVGPAPRAGVAGASALKURLASLATSLEH
RMLDVTADVADAVTVLVASPLDIEPGRPIQELGIDSLMAIELRRLAAATGLSATL
LPHPTPEALAKITSGILHDEAPAPSTSAEIDRLGALASAVANBAIRDELTVRMK
ALISKMAAEPSDDGLASRLDTANDLELFMRIDHVALEAAT"
22592..33133
/gene="sp11"
22592..33133
/gene="sp11"
/function="biosynthesis of spirangiene"
/codon_start=1
/transl_table=1
/product="polyketide synthase"
/protein_id="Cad3450.1"
/db_xref="GI:27528336"
/translation="MSPAPGTEALRDYLTQKLTDLORSOTRLHAESKAYEPIAIV
MSCRFPGASPELMQLEGGDVAVSGPAPSRMSVEPRVDPAVDTGKPLAEGGF
LEAGLELDAAPFGVSPREALEMPQOQLLETTEAEILAEITDASISQSGTGVGV
MNNYASHHIIGDSDAGSYSPASSAVASGRIVYTTGEPPTVDTGACSSVAVAH
LAQALSGSCALALAGVTVMTAREFDALSFSGSDGCKSSAANGGWSBEGAG
MVLIERASONGHPIPLAVIRGSVAODKSOQLTPNGAQDRVIRLOALAAARLET
HDVLEAVEHGTGPIGPIEANALATVYGGAPERLPLMGLSISNVTGTOAAAGVG
IKITVLAIRNGILPRTIHAHAPSHVDMAGGKRLADAPWTRNGVRVRAVSPGL
SGTHAVIYEAERPRPDEASPAESGDQALPILSKTEDALBAQAAARLHAGDP
GASLVDAASLATTIRAFDHRAPVANSRDLAALDAFAPAGLDAAPASVAVARRAP

```

Query Match 20.1%; Score 859; DB 1; Length 48000;
 Beet Local Similarity 57.6%; Pred. No. 7.7e-85;
 Matches 1735; Conservative 0; Mismatches 1220; Indels 48; Gaps 9;

Qy	14	CCATCGAGCGCGGCAAGCGGAAGATCGATTGCGATGTCGAGCGAGTTCGCTGTCGCCG	73
Db	12286	CCCCCGCGGACGCGAGCGACGAGCCATCCGATGTCGCGGTGCGCTCGCTCCAG	12345
Qy	74	GTGCGCGATCGATCGAGCGGGTTCGAGACGCTCTCGAGGGGCTCGCGGACACCGCGT	133
Db	12346	GCGGGGTGCGACGACGTGACGCGCTCTCCGCTGTCGAGGGGAGGCGACGCGGGTGG	12405
Qy	134	GGCGAGTCCCGCGCAACGCTGGAGTACAGACGCGTGTGATCCGACCCCGATGCC	193
Db	12406	CTCCGATCCCGCGGACGCGCTGCGGACGTCGCGCGCGGTTCGATCCGACCCCGAAGGA	12465
Qy	194	CGGGAAGACGCCCGCTTACGCGCGCATCTTCTCTGACGACGTAGCTGCTTGCACGCT	253
Db	12466	AGGGGAAGAGTACGTCGCGCACCGCGCATGCTCGACCGGTGTCGACCTCTTTCACGCGG	12525
Qy	254	CGTCTTCGCGATCGCTCGCTCGGAGCGCTGCGGAGTGAACCTCGACATCGATCTTGC	313
Db	12526	GCTTCTTCGCGATCGCTCGGAGCGGAGCGAGCGATCGACCCCGACGACGCGCTCTTGC	12585
Qy	314	TGAGAGTGTCTGCGGAGCGCTCGGAGACGCGCGATCGCTCATCGGCGCTCTCGGTA	373

Db 12586 TCGAGACCGCGTGGCAAGCGCTCGAGCGGCGGACGATCGTCCCGGCTGCTCCGCACT 12645
Qy 374 CGGAAACGGAGTGTTCATCGGAGATCGGCCCGTCCGAATATAGAGCGCGCTCCGCAAG 433
Db 12646 CGAACACCGCGCTTTCGTGCGCGCGGCGCGAGCGACTACGCG-----TGCTCC 12696
Qy 434 CGACGCGCTCGGACAGATTCAGCTCATATGGCGGCTGGGAGCATGGCCCAAGCTTCGAG 493
Db 12697 AGAGACGCGCCGAGGACGCGGAGGCTTACGCGGCGATGGGACCGCGGCTTGTTTCGCG 12756
Qy 494 CGGCGCGAATCTGTATGCGCTTCGAGGCTCGAGGCGGCTGTGCGGAGTACGAGCT 553
Db 12757 CAGGTCGCGCTGGCGTTACAGCTTCGCGCTCCAGGCGCGCGCTTCGTGATCGACCGCT 12816
Qy 554 ATTGCTCTGCGTGGTGGCGTTTCAATCGGCTGTGAGCTTGGCTTCGCGGAGATGCT 613
Db 12817 GCTCGCTGCTGCTGCGCTCGCTCACTCGCTGCGAGTGGCTGCGGCGAGGAGTGG 12876
Qy 614 CCAACGCGCTGCTGCTGGGAGTATGCTGATGTTGTTCGCGAGCACCTTCGTGGCTT 673
Db 12877 ACCTGCGCTGCGCGCGCGCGCTGCAAGTATGAGCTTCGCGCGAGGCTTTCGTATCT 12936
Qy 674 CGAAGACCGGCGCGCTGGCGCAGGAGCGGTGCTGCAAGGCAATTTTCGCGAGCGCAATG 733
Db 12937 CGCGACGCGCGCGCTTCGCTCGGAGCGGCGATTCGAAGCGTTTCGCGCAACGCGAG 12996
Qy 734 GGTTCGAGCAGAGCGAAGGAGTGGCGCTGCTGCTGCTTCAAGCGCGCTCAGTGGAGCCGCG 793
Db 12997 GGTTCGCGCGCGCGCGCAAGCGCTGCTGCTGCTGCTGCGCGCTTCGCGAGCGCGCG 13056
Qy 794 CGGACGGGAGTGGGATATGGGAGTATTCGAGGATCCGCGATTCATCGACGAGTGGCA 853
Db 13057 CCAACGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13116
Qy 854 GCAGCGGTCTGAGCTGCGGAGCGGAGCTCCCAAGAAATCGTGCAGAAACGCGCGCTGG 913
Db 13117 CGAGCGGATCATCCGCGCGGAGCGGCTCTCGGAGAGAAAGTCTTCGCGCGCGCGCTCG 13176
Qy 914 CGGACGAGAGTGGCGCGCGCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 973
Db 13177 ACGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13236
Qy 974 CGCTGTGTGACCCCATTCGAAATTCGAAGCTTCGAATGCGGATTAACGCGCTCGGCGAGATG 1033
Db 13237 CCTGCGGCGATCCATCGAGGTTCAGGCTTCGCGCGCGCTTCATCGCGAGAGGCGCGCG 13296
Qy 1034 TCGCACGCGCGCTGCTGATCGGAGTGGAGTGAAGCAACCACTTGGGCACTCTGAGTATGCT 1093
Db 13297 CGGACGCGCGCTTCGCGCTCGCGCGGAGTGAAGCAACCACTTCGCGAGTTCGCT 13356
Qy 1094 CGGAGATCATCTGGGCTGCTGAGAGTGTCTTGTCCCTTCAGCAAGCGGAGATTCCTGCG 1153
Db 13357 CGGCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 13416
Qy 1154 ACCTCAGCGGAGCGCTGAGACCCCGGATTCATGGGAGTATCTTGGCTGAGACCTGCA 1213
Db 13417 CGATCCACCAACCGCGCAACCGCGCATGACTGGGAGCGCTCCCGCTCGCGCTG 13476
Qy 1214 CGCGCGCGCGGACACCGTGGCG-----GACTGGAATACGCGCGAGCGGCGGAGTGA 1267
Db 13477 TCGAGCGCGCGCGAGCGCTGGAGCGGAGCGCGAGCTTCGAGAGCGCGCGGAGTCT 13536
Qy 1268 GCTGCTTGGCATGAGCGGAGCAACGCGCACGTGGTGTGGAAGAGCGCG-----CG 1323
Db 13537 CGCGCTTGGGCTTCGCGGAGCAAGAGCGCATGCTGCAAGAGGCGCGCGCAACCG 13596
Qy 1324 GCGAGCTGACACCGCGCGCGCGCGAGCGACCGGAGAGCTG-----TGCTGCTGT 1375
Db 13597 ACGCCCGAGAGCGCGCGCGCGCTGCTGCTGCGGCGCGCTGCTGCTGCTGCTGCTGCT 13656
Qy 1376 CGGCAAGAGCGCGTCAAGCTTGAATGACAGCGCGCGCGCTGCGCGAGCACTTGAAGA 1435

Db 13657 CGGCGAAGGAGCGCGCGCGCTTCGCGCGGCAAGCGCGCGCGCTCCGCGATGACTCGGCA 13716
Qy 1436 CTTACCTTTCGAGTGTTCGAGCGATGTGGCGTTTCACTGTGCGTCAACGCGGACGCGCA 1495
Db 13717 GGAACACCGGAGCGCGCTGCTGCAAGTGTGCGCGCGCGCTCGCAAGCGCGCTTCGCAAT 13776
Qy 1496 TGGAGCACCGGCTCGGCTGGGCGGAGCGTGAAGGAGAGGAGTGGGAGCGCGCTGAGAG 1555
Db 13777 TTGACCTACCGCGCGGATTCGCGCGAGCATCAAGGAGCGCTGTGCTGCGCGCTGAGAG 13836
Qy 1556 CTGCGCGCAGGAGCAGACGTTCGCGCGGCTGCGAGTATCGCGGATTCCTTCAACGCG 1615
Db 13837 GCGTGGCGCGCGCGCTGCGCGCGCGCGAGCGTT-----GCTGCGCAAGGCGCGCGCG 13890
Qy 1616 GCAAGCTGCGCTTTCCTTTCACCGGACAGGAGGCGGCAACGCTGGGAGTGGGCGCTGGCG 1675
Db 13891 ACAAGCTGCGCTTCTGTTCAAGGCGCGGCGCGCACGCGCTGCGATGGGCGCGCGCG 13950
Qy 1676 TGTAGATGTAATGCTCGCGCTTCGCGGAGGCTTCACTGTGCGTGAAGGCTGTTCAAC 1735
Db 13951 TGTAGAGCGCGTCCCGCTTCGCGAGCGGCTGAGATGCGGTGCGAGCGCACTTCAATC 14010
Qy 1736 AGAGCTGACCGCGCGCTTCGCGAGGATGTGGGCGGAAACCGGCGACGCTGACGCGCG 1795
Db 14011 GCGAGCTGATGCGCGCTTCGCGGAGCGTCTTCGCGCGGAGGCTTCGAGGCTGCGCGA 14070
Qy 1796 CGCTGCTGACAGACGCTTCACCGAGCGGCGCTTTCACCTTTCGATATGAGCTTCG 1855
Db 14071 GCTTCCTGAGCGGACCGAGTTCAGCGCGCGCGCTTTCGCTTCGAGGCTGCT 14130
Qy 1856 CGGCGCTGAGCGGCTGAGGAGTGAAGCGGAGTGTGCGCGGAGTGAAGTATGAGTATG 1915
Db 14131 TCGGCTGTGAGAGCGTGGGAGCTGAGCGCGCTGAGCGCGCGAGTCTCTGGGAGATTCGCGCG 14190
Qy 1916 AGCTGTGAGCTGCTGCTGCGCGCGCGCTTCTGCTTGAAGAGCGCGGTTCTGTGTG 1975
Db 14191 AGCTGCGCGCGCGCGAGTTCGCGCGCGCGCTCTGCTGCTGCGCGAGCGCTGCTGCTG 14250
Qy 1976 CTGCGCGCGCGCGCTGATGCAAGGCTTCGCGCG-----CGGCGGCGATGTGTGATTCG 2032
Db 14251 CGGCGCGCGCGCTTCAAGAGCGCTGCGCGCGCGCGAGCGCGCGAGTGTCAAGGCTG 14310
Qy 2033 AGGCGCGGAGCGGATGAGCTGCTGCGGCTGCGCGCGCGCGAGCGAGCGCTGCTGATGCT 2092
Db 14311 AGCGACGAGGCGGAGGCTGCTGCGCGCTGCGAGGCGGAGCGGCGCGCGAGATTCG 14370
Qy 2093 CGCGCTCAACGCTTCGAGCAGAGTGTATTCGCGCGCGCGCGAGCACCGCTGATGCGA 2152
Db 14371 CGGCGATCAAGCGCGCGCTTCGAGCGGATTCGCGCGGAGCGCTGCGCTGCGCG 14430
Qy 2153 TCGCGCGCGAGTGGCGCGCGCGCGGCGGAGCGAGCGAGCGCTTCCAGCTTCGATGCTG 2212
Db 14431 TCGCGCGCACTTTCAGACGCGCGCGCGGCGAGAGCGAGCGCGCTTCGCTGAGCGCGCT 14490
Qy 2213 TCCACTCAACGCTTTCAGCGCGCGGATGTGAGAGCGCTTGGGCGTGTGCGAGTGTGATG 2272
Db 14491 TCCACTCAACGAGTGAAGCGCGCTTCGAGCGGCTTCGCGCGCTTCGAGAGGAGCTCA 14550
Qy 2273 GCTACCGGCGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2332
Db 14551 GCTTCATTCGCGCGCGAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 14610
Qy 2333 AGGTAGCTTCGCGCGGCTTATTTGGGTCGCGCGCGAGAGTGTGGCTTCGCGGATG 2392
Db 14611 AGATCGGCTGCGCGCATCTAGGCTTCGCGCGCGAGCGCGCTTCGCTTCGCGAG 14670
Qy 2393 GAGTGAAGCGCTGACCGCGCGGCTGCGGCGACCTTGTGAGGCTGCTGCAATATGA 2452
Db 14671 GCATCAATAGCTGAGAGCGCGAGCGCTTTCGCTTTCGAGCTTCGCGCGCGAGCG 14730
Qy 2453 CGCTGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2506
Db 14731 TCTCTCGCGCGCTGCGGAGCGCGCTGCGAGCGCGAGCGCGAGGCTGCTGCTGCTGCT 14790


```
/protein_id="AAK57185.1"
/db_xref="GI:14210838"
/translation="WRAGGGPHLLGVSTTSVQKENTFWESALSKEPSYADHRVG
GAVVPGAAVYEMALSAARELSGEKAYELVDTSFKEGLLPEQGERVVMALSGEGD
ALSPOLSKASTENPAAGMAHAYVGOIRPEGSPQTQPSFPAIRERCEVLAQA
AHYELASRGVAYPGPOGOOIMRGRAELGHLPEPLARASAVIRHPALNACF
OLVAASVPOEDLTPDAGPAVVALAHLRLHEDPGTEVPCHTVALRPSBGRVYEDL
VLRDASGRILVEALRVOQLDPAAHAGDLPALGERHNASVQGPAPBPGASGR
WLLADSELDATVESQLEKREGEVIRVDLAGAGGRVVDATSPQAFGVLSAEGA
GSAPRGVHLTGLEPREVDPIMGCGVTLVQALSRSVLASPRMLVTRGVLSAEGS
RSADVQASLWGLGRTLAHSEELKCTRVDSAMGTGDAGALMRELVADDEEIS
SLRADRGVIRGSPGRAPSEAVPAGRGEPFLERDQGLEFRAGARPGGEL
ELREVALRGEGCGEPMEGAGEGICGSRVVAIGGVVAVGGERIALVAGSL
GSHVVPACCTAIRPSSSHGAAALATVPAVALPHLGLERKGVWVGLATSLG
GMAIKLARAGAEVFAVATDEQASLRELDQVMDPSDFSVRLDADTGGGVE
LAIVAPGVELERSLPTLSEGRVLDLRASGASARTGDANVAMCVVDQDPARRRER
FASLQVEVAAFEAGGLPSAADARALAKDPKARVVVPAADARLRADGLVTVGLG
GGGLAVKAMVEOGAGHLVLVGRDVTVPQOEVVTVLEAGARVPAADAVDRAO
IARITSEIAEGGAPLRGVTHAGVLDGVLLOQTVRPRVMAPVVIGGMNHLVTRRE
APLDPFVLYSSAASMPGAPGCGNYVAGNAPFDALAHHRALGPELSTWGPFSVGL
AAQANRGRRLAQSGDSITTPAESNTILGRLLDGVTOAWVPLELRKWEYFPKAS
SPWLSLVPADAGVEGEARDVALLETLTATPOAKALGFVREQLARLUDSAR
IDPEAPLOSFGIDSLMGLERNRLVSGSLPASIWKHPITLALCAHLGEVMDRT
LAETLAQAAAGPAPDESADKXEVTL"
complement(11487..14303)
/gene="mxab2"
complement(11487..14303)
/gene="mxab2"
/translation="MSKOEKPAEPSPALORALALVERKMOARLDAVERARTEPIAVIGL
GCRPFGAVDSSYWKILRDGDVALREVDESMDVPGHDSRQVADPACRGGFELD
DYEHDAERFGISPREAASLDPOQLVLEVAEMALBNGLADQLAGSKTQYPMKVS
SDYMRLEKENDATRFDMATGNGSFYPGRVSYVLGQCGKMPVDTRCSVALH
LASERLRAGESNLALAGVNLILSEPMLCLCSMOALASDGCKFDASADSVYEG
GALVILKRLSDAQRDGDSLIALIRGSAVHHDGASGLTVPNPNSQOAVORALDNARI
APALVYIEAHGTGTPLGDPLELRALGAVLIGKRPADRPFFIGSVKTNIGLEPARI
AGVITITLISLHKEIIPHLHPRTNPHVEMDRIIPARVPERPWPAGRRIRLAKGSE
GLSGINAVVLEEARASTPTVAEDRAQLVLSAKTPRALSIAZMSSTLAKBEG
GSLADISTLSLSTPHEHRLALAKSKOELIEOLEAPRAGSHAGLSGKRGTLGQ
KIVFVPGQSGQMLMGQRLVAENVTFERAILDWCHEALSRYADWLRLDITNPQOPI
MNDIVIDVIOPTLFALQVLAAMVRLGLEPHEVAHGHSMGVSVAALHAGALSDDARI
CERSKLRRVSGOGAMAVVDSRBOAEVIRGAEPRIALAVNGSRSTVLSGDPALAK
EVLRLERREVECRWVKVDVASHSPOMPLRGDLDPMALIRPTGSAVLYSTVGAP
IDGRRLSAPYWRNLRQPLPADVORLIDQSNVLFELSPHPIIIVPVDMALREGDT
SGRGLVPSLIRHDEMPSLSSVQALFTFVIAMTSGITPTSVAGLRFKTRGSASG
TGMSCQQLRRRGSCALAREVIPS"
complement(14300..29878)
/gene="mxac"
complement(14300..29878)
/gene="mxac"
/translation="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
ACP regions"
/codon_start=1
/translation="mxac"
/product="mxac"
/db_xref="GI:14210840"
/protein_id="AAK57187.1"
/translation="MANRDESNELSPAKRMVALEKQTRINAVEGAKEPIALVGM
ACRPOADNVEAFMRLLQGVDAVEIPIELERPKQEVLDIGGARNGWGLDVSVPD
PGFPGISPREAVRMDPOQLLEVAEMLEADGLVDKASAEVLFIGA CDFDSVDF
OYERPETGAFSAGTAVASVYSGRLSYENLGPSLVYDTAGSSSLVLAHCOGLNR
BCRNALAGCVNLILSPSVLSKRLGSLPGRKAPDASNGSTRBEGCGVILKRL
ISDALADGHIILATTRGSAINDGKSTGLTTPNVLSQALIKQALESSGLKAEQVSY
EAHGCTGPLGDPLEALRETYGVPRPDSVCGISVKTNGVHLSAAGIAMKIVL
AMRHOSIPHLIKQVNPRIQLEGSLTITPRLTPMKASGQPRRAGVSGFSGISNAH
VIEBAPAVAKDALPQREVLPLSARSEVALALERAERVAELGEGGSLSDCYTA
SAHRTHTHRLVLADREHTERLRAFAFRBQSPGASGRAPFGRRKILVYVYVQCG
```

```

AORAGMBALMESEVPFDALVKVDKALQPLHGLMSVLDEIADREBSRLAEDVNOVP
LPAYVLTALMRSMGIEBDAVMHSMGEVVAAYAGSLIEDVLTICHRSKIMRTL
GGRGTAMVNLVSEBESAIRPNMGKIULGGLNRPQSVISGEPVAKVEALNGKR
GVFRVAVMDVPSHTPLVEVLSRELEGIATITPRAIDPLCSTVTGFFLGAKIDAA
HMGNLRNPVPAALAOHLILASDGTPEVSPHVLPAVQTMASGTEVLAASLR
RNEBDRHTMLSLSALYSSGRVAMNRYPEPKGQVLPPTVPMQERFMBEPVAVGR
ASRSQSGHPLGERILDLSAGEVRLMOVQGRPHFLAHRVQGVLPALPAANAD
ALSAKQITGBPVSLDEVSFSAISPEDEERYOLMISTEPGASRLFSAPMRE
TEBDELTAHABESLRHGDASVYKAPVALDELARCSEASVEAHYQTLKEIGLE
YGPFRVAVRELIRKAGELGVOLLEGELTAEERSLHPALDQFOVLAALSGDS
EGGVVVPVALSRLVLVPEBALMSHRIVSGAAGSNTIEDVTLIDQORVAEVR
RLVCKILAAASAPADEVDRIYSAQMEPOQOAALETSSRTAGRMVLFAARIGVS
LIGLAEABAGCVVVEAGVSYREABDRYNPSKLEPFOALKEAFDQORPCGMVH
LYGLDVASNTPEPSFREALLSSTSVYILOANQORMDTPMLMVLTVTSVOYASG
EOPSPAAPIVGLAAVITTHPELRCHVDLSNOPSEBELASLTLEELSGABRILA
LRGAVRHVARELRKISGKAAEPREFTADGYLLTGLGAGGLEVAKFETORGANILA
LTGGGAPBEEQKTVLELRNAGAEVRVRAVNTSAADVARLLGSDLSPLRGVHA
AGVADGLVTLDEPRLRAVLA PVQGMNHLVOTQGRPLFLFVMSAASMGSPGO
GNVAAANAFMDVLAHHRPAQLPGLSVVMGPMADVGLAAVATNMGDRALRGVMP
AQLAEALGRLLVQBAAOVAVVPLDRLQWREERYLSAOSPFVLSVMOEKVSTPASRKS
ARELLAAAEPRQRKILEGVYREOLGRILRNEPRRIEDQPFQDLGMSLTLGELRNR

Query Match
Best Local Similarity 58.4%; DB 1; Length 49736;
Matches 1563; Conservative 0; Mismatches 1091; Indels 24; Gaps 4;

1 GTGGCGGATCGTCCATTCGAGCGCCGACCGCAAGATCCGATTGCGATCTGCGAGCGGCT
19792 GAGGAGGAGACCGCGTGTGGGCAACCCACGAGCGGATTCGATGCGGCAATTGGGC
19792 TGCCGTCGCGCGGAGCGGTATCGATTCGAGCGGGTTCGGAACGCTCTCGAGGGCTCG
61 TGCCGTCGCGCGGAGCGGTATCGATTCGAGCGGGTTCGGAACGCTCTCGAGGGCTCG
120
19792 TGCCGATCCCGCAGGCTTCG---GTCGCGAGGCGTTCTGGAACGCTTGGAGGAGGAC
19675 GTCGAGCGCATCCGGAAGTGCTGCTCGCGGTCGAGAGTCGAGGACTACGACCCG
180
121 CGCGACACCGTCGCGGCGAGTCCCGCGCAACGCTGGAGTGAAGCAGCGTGGTTGATCC
180
19675 GTGAGCGCATCCGGAAGTGCTGCTCGCGGTCGAGAGTCGAGGACTACGACCCG
19616
181 GATCCGATGCCCGCGGGAAGACCGCCGTACGCGCGCATCTTTCTTGAAGGAGCTAGCC
240
19615 CAGCCCGCGTCCCGGGAAGACCTACACCGCGTGGGTGGTTCTTGAATGGGCTAC
19556
241 TGCTTGAAGCGCTCTTTCGCGCATTCGCGCTCGCAAGCGCTCGGATGAGACCTGCA
300
19555 CAGTTTGAAGCGCTCTTTCGCGCATTCGCGCTCGCAAGCGCTCGGATGAGACCTGCA
19496
301 CATCGACTCTTCTGAGAGTGTGCTGGAGGCGCTGAGAAACGCGCGCATGCTTCATCG
360
19495 CAGGCGCTCTGCTGAGAGTGTGCTGGAGGCGCTGAGAAACGCGCGCATGCTTCATCG
19436
361 GCGCTCTGCTGAGAGTGTGCTGGAGGCGCTGAGAAACGCGCGCATGCTTCATCG
420
19435 CGCCTGGCAACACCGGCACTGGCGTGTTCATGATATGGCAAGACGATGACCGCATG
19376
421 GCGCTGCGCAAGCGCGCTCGCAGAGATCGACGCTCACTGCGCGGCTGGGAGCATG
480
19375 CTCACGCGGAGTGGGAGCGCTCGCGCGCAGAGGATGCTTCATGCTCCACGCGCATGAC
19316
481 CCCAGCGTCGAGAGCGCGCGCATCTGTATGCTCTCGGCGCTCGAGAGCGCTGTGTCCG
540
19315 TCGACCTTTCGCGCGCGCGCTGATGCTGCTCGGCTTCAGAGGGCGCAGATGAGC
19256
541 GTGATTAAGGCTATTCCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
600
19255 TTGAACACCGGTGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
19196
601 TCCGCGGAATGCTTCACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
660
19195 GCCGATGAGACCACTTGGCGCTGCGGCTGCGGCTGCAACATGACGCTGCGCGGCACTCG
19136
661 CTCGTGTGCTCTCGAGAACCGCGGCGCTGCGCAAGGACGCTGCTGCAAGGCACTTTTCG
720
19135 ACCATCTACCTTCGCGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
19076
```


Qy 721 GCGAGGCGGATGGGTTTCGACGAGGCGAAGGGTGGCCGCTGTCGATCTTCAAGCGGCTC 780
Db 19075 GCTCCCGCGAGCGCTACGTCGCGAGGAGGGTGGCGGATCGATGTTCTCAAGCGGCTG 19016
Qy 781 AGTGAAGCCCCCGGCGGACGCGGATCGGATATTGGCGGATGATCGAGATCCGCCATCAAT 840
Db 19015 TCCCATCGAAACGGGATGGGGATGACGTCCTGGCCGCTCATCCGCGCTCGCGCATC 18956
Qy 841 CACGACGGTGGGACGACGCGTCTGACCGTGCAGAACGGGACCTCCAAAGAAATGGTGTG 900
Db 18955 CACGATGGTCCGAGCAGTGGCTTCAAGGTGCTCAATGCGATGCCACACAGATTTCTC 18896
Qy 901 AAAGGGGCGCTTGGCGGAGCGACGCTGCGCCGCTTTCGATGGGTTTATGTCAGAGCAC 960
Db 18895 CGGCGGCGCTTGGCGGAGCGGAGGTTGGCGCTGCGAGCGTGGATTACATGAGGCGCAT 18836
Qy 961 GGCACGGGCGACGACGCTTGGTGAACCCATCGAAATCCAGCTCTGATGCGGATACGGC 1020
Db 18835 GGAACGGGCGACGTCATCGGCGATCCCATCGAGGCTCGAGCTCGGCGCAAGTCTCGGC 18776
Qy 1021 CTGCGGCGAGATGTCGCGACGCGCTGCTGATCGGGTGGTGAAGACCAACCTTGGCGAT 1080
Db 18775 CAGGACGTAAGTCCGAAACAGCGGCTGCTGATCGGCTGGTGAAGACCAACATTTGTCAC 18716
Qy 1081 CTTGATATGCGTGGGAGTCACTGGGCTGCTGAAGTCTGCTTGTCTTCCCTTACAGCGGG 1140
Db 18715 CTGAGGAGCGGCGGCGGATCGCGGATCTCATCAAGGTCGTCCTCGCCATCGGCGATGGG 18656
Qy 1141 CAGATTCTTGCGACCTTCACGCGCAAGCGCTGAACCCCGGATCTCATGGGATGATCTT 1200
Db 18655 GTGGTTCCGCGCACCTGTCACCTCAAGAAAGCTCAACCCGACATGACATGAGCTTC 18596
Qy 1201 CGGCTGACCGTCAAGCGCGCGCCGACACCGTGGCGGACCTGAATTAAGCGCGAGCGG 1260
Db 18595 CCGCTGACATCCGACACAGGTCATCTCGTGGCGGAGCGCTGCGGCTCTCGGATGCT 18536
Qy 1261 GGGGTGAGCTCTGTCGCGATGAGCGGAGCAACCGCGACCTGATGCTGAAGAGCGCG 1320
Db 18535 GGCATCAGCGCTTGTGGCTTGAAGCGGACCAATGGGACGTCCTGCTGGAAGAGGCGCC 18476
Qy 1321 GCGGCGACGTGACACCGCGCGCGCGCGGACCGGACGAGCTGCTGCTGCTGCGCA 1380
Db 18475 GCTCCCGAGCTCGCGCGGAAACAGGCGGAGCGGAGCTGACGTCGCTTGCCTCGCG 18416
Qy 1381 AGGACCGGTCAGCGCTGAGATGACACAGCGCGCGCGCTGCGGACCATTTGGAACCTAC 1440
Db 18415 AAGACGAAAGGTGGCTTCAAGCACTCGCGCTCGGCTGAGATACCTTGGCGAGCG 18356
Qy 1441 CTTTCCGAGTGTGCGGATGATGCGCTTCAATCTGCGGACGACGCGGACGATGGAG 1500
Db 18355 CCAAGACGAAAGCTTCGCGGATGTCGCGCATGCGGAGAACGACGACCTTGGTTCG 18296
Qy 1501 CACCGGCTCGCGGTCGCGGACGTCGAGGAGGGGCTGCGGACCGCTTGAAGCGTGGC 1560
Db 18295 CATGACCTCGCGCTGTCGATGAGATCGGATGAGACGACCCGAGCGGCTTGGCGCTT 18236
Qy 1561 GCGGAGGACGAGTGGCGCGCGGTGGCGTGGCAGTATCGCCGATCTTCAACGCGGACG 1620
Db 18235 GCGGAGGAGGAGATGTCCTCGGCTGGCCCAAGACCGCTGCAAGGAGGATCCCAAG 18176
Qy 1621 CTCGCTTTCTCTTCAACGCGACAGGGGCGAGACGCTGCGCATGCGCGCTGATAC 1680
Db 18175 GTCTCTTTCTGATTCAGGGGCGAGGCGCGACGATACGCGGCGCATGCGCGGAGCTGAC 18116
Qy 1681 GATGTAATGTCGCGGCTTCCGAGGCGTTCGACCTGTCGTAAGGCTGTTCAACGAGAG 1740
Db 18115 GAGACCGAGCTGTCCTTCGCGCGCTTGGATCGGTGCAATGCAATTCAGAGGCGCAG 18056
Qy 1741 CTGACCGGCGGCTCCGCGAGATGATGGGCGCAACCGGCGACGTCGAGCGCGCTG 1800
Db 18055 TGGGCGAGTCCCTCTCTCGGATGTC-----ACGGCAGCGGCTCGCGG 18011

Qy 1801 CTGACGACGACAGGCTTACCCAGCCGCGCTTTCACCTTCCAAATATGCGCTCGCGC 1860
Db 18010 ATTGACGACACCGAGTACACCCAGCCCGCTTGTCCGTTGGAGTACGCGCTCGCGAG 17951
Qy 1861 CTGTCGCGGTCGTCGCGGATGAGCCGAGTTGGCTGCGCGGCACTACATTCGAGCTG 1920
Db 17950 CTGTCGCGGTCGTCGCGGATGAGCCCTGAGCGGCTGTTGGGACACAGCTGGATGAC 17891
Qy 1921 GTGCTCTCTGCGGTCGCGGCGCTTTCGCTTGAAGACCGGATGTTCTGTCGTCGTCG 1980
Db 17890 GTGCGCGCTCTGCGGCGGAGTATGAGACTCGAAGAACCCCTCGGCTTCATACCGAG 17831
Qy 1981 CGCGGCGCTTATGACGAGCGCTGCGCGCGCGCGGCGAGTGTGTCATGACGCGCG 2040
Db 17830 CGTCGCGGCTCATGACGACGCTGCGCGCAATGGCAGATGGGCGCTTCCGCGAGT 17771
Qy 2041 GAGGCGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2100
Db 17770 GAGGCGAGTGCAGGAAATGTCGTCCTTATGCGGATGCGGTTTGTGATCGCGCGCTC 17711
Qy 2101 AACGCTCGGACGAGTGTATCGCGGCGCGCGGCAACCCGTCATGCGATCGCGCG 2160
Db 17710 AATGCGCGCGGACGACCGCTGATTTCCGTCGAGAGGAGCGGTCGAGACCATTCGCT 17651
Qy 2161 GCGATGCGCGCGCGCGCGCGCGGACCAAGGCGCTCCAGCTTTCGATGCGTTCAC 2220
Db 17650 TCGTTGACGCGCAAGGCGGTCGCAAGGCGCGCGCTTACGTCGTCACGCGCTTCACT 17591
Qy 2221 CCGCTCATGCGCGCGGATGTCGAGGCTTGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTC 2280
Db 17590 CCGCTCATGAGACCCGAGTCGAGTGTGTCGAGAGGCGGCTGCAAGATGCTTCCGT 17531
Qy 2281 CGGCGCTGATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 2337
Db 17530 GCGCTCGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 17471
Qy 2338 AGCTCGCGGCTTATGAGTGGCGGACGCGCGGAGAGGTCGCTTCCGCGATGAGTG 2397
Db 17470 CTGACGCGCGCATCTTCCGCGCGGACGTCGCGGAGCGGCTTCCAGAGCGGCTG 17411
Qy 2398 AAGGCGCTGACGCGCGGTCGCGGACCTTCTGTCGAGTTCGTCGGAATTCGACCTG 2457
Db 17410 AAGGCACTTGGGCGCTGCGTGGGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 17351
Qy 2458 CTGCGCTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2517
Db 17350 ACAGCATGCGCTGAGAGGCGCTGAGAGTGGGACCTTGTACCTTCCCTGCGC 17291
Qy 2518 GTCGCGCTGACGAGCCGCGACCGCTGCTGAGCGCTGCGCGGCTTGGGCGCTGCT 2577
Db 17290 AAGGTCAGGCGCGAGCGCGGAGATCTTTCAGCTTGGGCGGACCTACACCGAGG 17231
Qy 2578 GGCCTGTCCTGCGGCGCGCTTCTTCCCTCAAG---GGGCGCGGAGTCCGCTGCC 2634
Db 17230 GTGCGGTCGAGATGAGTGGCTGAGACGCGAGCGGACCGGTCGCGCTGCTGCTCCT 17171
Qy 2635 ACGTACCTTGGACGCGGCGGCTACTGATGTCGAC 2672
Db 17170 GTTACCGGTGACGACGAGCGGACCTGCTGGAAC 17133

RESULT 14

SAU421825

56808 bp DNA linear BCT 16-MAR-2002

LOCUS

SAU421825

DEFINITION

Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,

KEYWORDS

acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;

kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;

ORF7; ORF8; ORF9; polynucleotide adenylyltransferase; potassium

SOURCE	channel beta chain; ribosome binding factor; stia gene; Stia protein; stib gene; Stib protein; stic gene; Stic protein; stid gene; Stid protein; stih gene; Stih protein; stif gene; Stif protein; stig gene; Stig protein; stih gene; Stih protein; stij gene; Stij protein; stik gene; stil gene.
ORGANISM	Stigmella aurantiaca
REFERENCE	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales; Cyctobacterineae; Cyctobacteraceae; Stigmella.
AUTHORS	1 Galatzis,N., Sliakowski,B., Kunze,B., Nordstiek,G., Blocker,H., Hofte,G. and Muller.R.
TITLE	The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmellin is directed by a novel type of modular polyketide synthase
JOURNAL	Online Publication
REMARK	J. Biol. Chem., 10.1074/jbc.M11738200
AUTHORS	2 (bases 1 to 66808)
TITLE	Muller.R.
JOURNAL	Direct Submission
REMARK	Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124 Braunschweig, GERMANY
FEATURES	Location/Qualifiers
source	1..66808
	/organism="Stigmella aurantiaca"
	/mol_type="genomic DNA"
	/strain="Sg a15"
	/db_xref="taxon:41"
	complement(4978..6231)
	/note="ORF8"
	/codon_start=1
	/transl_table=11
	/product="cytochrome P450 monooxygenase"
	/protein_id="CAD19077.1"
	/db_xref="GI:19572310"
	/translation="MRSVSASRPPSPANPAFLARLDAFYADPPFLYAMERD PRRMPDGLFLTRWADLCYRDRTRSSDGRARGAQYGDAPLFEHHTSLVNDP PLTRVRRLIVGLALTPRALSTPERGLTVLVLGLAVKAGADLIEDPAATPIEVI GNLADIPEERGPLRGSALITGALEPRITTAQECGNEAVTEFLDYLRKIVAGRRAR RDDPATDVLTRLIQESDDESLTELHQCVFLNAGHETTNLIGNALELAFPP EBARLLARAPALPTAVEEVLRYESSNQLNRRAADTEIGVAVPTGTFELTICAAH RPARFDEHEHVDGROPBRHLAPAGANTCAGMLARIEARIALAFLARPPDYALT APVARARARFGFTAMPTRIGSLR"
	6369..7730
	/note="ORF7"
	/codon_start=1
	/transl_table=11
	/product="protein kinase"
	/protein_id="CAD19078.1"
	/db_xref="GI:19572311"
	/translation="MBELISEALPDGSRVGPRIIGRAGYTYGAVTRAHVGRICPR SVALKLARYPNLRFERESELLERVHANVPRLLGHGWRAPOGEAHYVMQWEG VLYQMNKERLTAREVLRLAQNARALATMRGILHRDVGENVLSPBEHSLMD PCGTWEGEALPLTGGILAPGTKAYSPEALRPHMNRNEADAHYQATPDVYALGVT AVRLCGTYPPTATPSGSGEDERETQEKRPQGVRLVPALBDLILBMLDNKER GRAGELIAEMELARBAQEADDEPNRPHSSVPSEITGRFVSVQEGMGHPPALVVA VLLLMGSSLTGRVAPASRMTGGTGADVALEALPVSSAREPSPSLNDML KSLPGQRPPPCRFQIStGGCMLVPOSSPCGSYYEMWACYSVPAAPRRTS DPQ"
	complement(7849..8577)
	/note="ORF6"
	/codon_start=1
	/transl_table=11
	/product="methy1 transferase"
	/protein_id="CAD19079.1"
	/db_xref="GI:19572312"
	/translation="MSQNTYDDSPFESYSOLARSLKGLAGAPEMPALRALIPDYKL RYVDLCCGRGFCRNARKUDAHVYGLDISEKMLRAKETSGDAITQADLEHLYP PGAFDLAFSSLAIRHYVERLEPLATVHGLTPGGHFWSTIEHP1FMASROEMTGP EBRHTWPVDSYOLEGKRITLHMALEEVKKYHRTLGLTNALLBQGFLLRIHEWGPTDA QLSAHPLEAEDRPMFLVSAQR"
	complement(8622..8879)
	/note="ORF5"
	/codon_start=1
	/transl_table=11
	/product="acyl-CoA binding protein"
	/protein_id="CAD19080.1"
	/db_xref="GI:19572313"
	/translation="MALEDPRSROERKALKTRPSNDRLLETLSLFRQATEGVQVK PGGLDYKERAKYAMWGRKLGRESMGOVVALVERLLRG"
	8961..10070
	/note="ORF4"
	/codon_start=1
	/transl_table=11
	/product="potassium channel beta chain"
	/protein_id="CAD19081.1"
	/db_xref="GI:19572314"
	/translation="WVYLCKSRGSQLKRLTNGSAHRRFRPMGCHARRPRMFRHGRS GLAWSEISYGNWLTHGSOVEAAAVACRALDGDITFTPDADVAGTRAEVIGRAL QGQRREGLEIPTKYWPTGPNDRGSRKRLIESIGSLRLKTEVVDLYOARVDT ETPLEETWOAFSDVROGKAHYIGVSEWADQLREGVLSROMGFOLVSSQPOYSLMW RVEBAVVPASOEIGVGQVYVSP1AOGVLTGKVRPGKPPGSGRATGKCSFTKRP LNDVLTFRVQDLTPVAQEGSLAQALAVANTLQNPNNSSALITGASRPEQVENYKAG VLSBEARHIDTLGLPIVERNPCKTLENSPRKP"
	10428..10862
	/note="ORF3"
	/codon_start=1
	/transl_table=11
	/product="ribosome binding factor"
	/protein_id="CAD19082.1"
	/db_xref="GI:19572315"
	/translation="MTSEFKORRRRAPRGTESSLSSEFAOSLPAMHLRVOSTLFOEVA RLFGELSDPLLEGVSVTSFELTPDGNARIGFTLPEYAGSGAP1EALVRSAGFI RAOLALGNLKRVRHLRFVCVGAAPGGS1PEDSSAEEGCS"
	10859..12016
	/note="ORF2"
	/codon_start=1
	/transl_table=11
	/product="cellulase"
	/protein_id="CAD19083.1"
	/db_xref="GI:19572316"
	/translation="MRPRIRIAGAGAVPIIAMARAVPAGAVKQOLAROPYAEHVA AMPDVHAGDVHNGRGSLSPGLDIGLSTHRLQREWERLVPHLCTLGGGNHLEL AFAAVPAGDVHNGRGSLSPGLDIGLSTHRLQREWERLVPHLCTLGGGNHLEL DRDLEGALMLLHTRGSGVGAIAAHMLRVAVQSGEGSLPOLRDTSGAACLADTEL ACQPARANRDLTORAVEVLAGAIGQGVLDSDRVADVHNHVAASHSFERTLLVIRKA VGLVAGQGGILRSGMGTASVYLVAKGDEQARCSHGAGRWTREARERTRPALAH ALRVVDPBGRVTLVERAPYARVDIEVELEDAELVTLPRRLPIAVLKG"
	complement(112082..11374)
	/note="ORF1"
	/codon_start=1
	/transl_table=11
	/product="cellulase"
	/protein_id="CAD19084.1"
	/db_xref="GI:19572317"
	/translation="WKSLLLLMLAVLPLVYTAALARPANABEGHFEGBENAGLTA AGCPRPVLPSTRGYIVDRCEGRFELSVNMFSGASDOLEVVGELDKOQLSDLVAGMKA MGFNSVRLPFPENLHPDNKPAQORCLEKTSVTCIDSTKNPDELNTKALVSVAGMVA ELTQOGLVLIINHTKSMCCMGDNGFNSQALODWODMTLAKRYQGNKVVA ADLENEVRPDGLDSBNMGMRHODMHAAMTGMGLLRTNDDLIVAEAVMMGLDGS SRPOLKPRQRPVALLRDKLYIYAHNTGYTGRPNQSGSLSSGKYSMDCTYXGTL DQEMGFALAAQATATAPWMSFEGVINEQANSRANFSNLADVLIEDVWATYALIN AAKLQNSVQGEDETYGLMSYPDMNGVRSGDRLTGRLGLGADGRGAVDATRFLP MTVLLKDGSSITYVDNNSLSDMSHGKAKLSCPGYRWGVSANFTILCTNAGAVPAQ OGTSYORTIIOEVSPILRYGDMANOSIKFECPTGVAVGISIAMPFLLEAGLCEON TCGTPIANKSAKNFMGQDQRASSSGGMSVSQYKQCQCDNXYLLGLARRSGLDYPS VVLCSN"
	14345..21466
	/gene="stia"
	14345..21466
	/gene="stia"
	/function="pks"
	/codon_start=1
	/transl_table=11
	/product="Stia protein"
	/protein_id="CAD19085.1"
	/db_xref="GI:19572318"

CDS

CDS

CDS

CDS

CDS

CDS

CDS

CDS

gene
CDS

/translaction="WTOLASRLGVGSESIDPRERFRYGDIDSRDTKSLIAELGALLKRL
 ESPTLMTYPTVEALSRLHAGEADVSGERADITGVNDPEPAIAGVAGFPAPAD
 BLSFSLMDVDAITEVPERMDDAFYSVGLKSGKGTGMDGLDVEDPDLFPFG
 ISPREAIOMDPOORLALBSWALESAVAUADTGMDSVFMGMMSDVAKLRGAL
 GLIPHTRGODSSIIARVSFGLRPSLTIVTRCSALVAHVALCQSLRNEBAM
 ALAGVNLILSPRESSVMSQJALSPDRCAFDKRNAGVYRGGEGVVLPLSRAT
 ANGYPVLCVIGSAVNDGFSNGLTAPNPQOEVLRAVARAGSPDKHYVEHTG
 GTVLDGPIEVALASVLGNRPAERPLIGSVKTVIHLBAKAGIAGMKVLIQHR
 AIPNLHGEONPYIRFEBRLVVKSLHPNNEGESLADGSSFGFTGCHVLA
 PPSGAHIAPIAETPEBLSILARQARAVVSGOVVALDCEPVARCEGSGORLA
 LIASPAELAAQVDFLAGERKPCILREBVTSGSQKVPYFSGTSGSPMGMBELR
 TEVPRGAIERDRLRGLSGMSMLEBSAGASRLGECGLQMLPQIOMVLAALM
 RSWGIEPVGWHSIGEVAAHVAIGLMBEARVITHLSLMGLSOGAKVILIDTS
 AEBANRLREWERHVVVAACNAPASTVYKAGEVIEBGLERGIYAHRNIDTPS
 HSPIMDPMEGLAEVLAGISPRVAVPVSTVTGFLAEECTSAWYWRNREPILDS
 QALTKFRODLRFLFELSPPHILARSVOCCKGROGVVALSMRGENBACRESL
 GRFLTGTSVAMQALYPAEAKRVELPMGLTGSIDLRGSRLLPLSAHTAEALEDA
 ARALQASTAKHPDRLDLCYSASVGSVHARILAVIDTSKALVEHLDDLKESRP
 GRINGCATGEPKRVFVFPQSGSQMGQQLRQEVFREAIEBCAALRSCAGWS
 LIDALQADPEHSLGSLIEVITPLFAIEVALAALMRSGVVPDVAVGHSMETAAVY
 AGALRTDAARVTCERSQLRRAVAGQGMALVELSLBOARMLSKDGRVSIAGNSP
 RSTVLSGDAIALBEILRELEAOGTCRCKVAVSHSPVDPLSDDLRVLSTAPRA
 ATPVLPXTVTPGEPIEGANMDASVYMRNIREVYLRQALIERLAGHGHVPLSHRPL
 LPSIDQTLRKHSGKGVLPRLREBEPKVALBESLGLTFTGGGRVDFRGLHPLGHRV
 RLPHLPWQRFLMDMDVAAGAASVGAAMPGGGSEHPLGHRPLSLAHAPCHVME
 NKLSRAVSYSLIDHVRGVTGVLPGTSYTEMALAAAKSGKTVLSLRVLEBKMSPL
 ESTASRTVSLIOEBAGHSFVLHGRSTPDEPTWLAGGTVHVEGDDEPPADPLMA
 AVEYRVLQKREVASDIAVEAMHAGNEVGPFGIEBOMVDAELVALRDEPLEOE
 LPEYRPHALDIDLOALIVAAEHGECGLAPRLESIRAHASGAMKMYNLPV
 REASFSGEVFLDESGORLAEQARQIIGESRPHDGLALMDLHEVMOPLPV
 VPTAPSSQGRWVLPADGTVAEGLAOLARGEICIVLBSMARFAEMEDHFPVDP
 QPEBVRKLFGLASTRTIPLRGIVHLMELDAPSPBEGCVLDSGMRVSGSVSLIQ
 GLNAAELAKQRLMLVTGVOVAGSAGCAVAOSPLMGQGVVAHYAHLRCTRAD
 LAPVSPBEEOTLLOELIAPAGDDOVALRGRHRYHPLRTRCTOVAVASRDLRAGI
 RSDCYLITGSLGIGLITAKMIVAKGRHIVYGRRAASMAEIALQMRGAGACEL
 VMKADVANTSVEBELVGTIQRKMPPLKRIFFAAGLADGTVALQLOARVMAKVE
 GAMNLRATILASLDFVLFSSASLVIPOGNYAAMNAFLDOLAGLRKRGACIOS
 IOMGPSEIIGLAEBQDNRGERLALRGMGLNPREBELMLLSGEBLRRARLIGV
 RLDRMRCKSYPAIRSSPLFSVILAGEENAOVERSEGLRLINDEPRTLRALIESF
 LREBAOYLRLPARIDLRSLOTIGMSSLILERNLLENLEITLDPATVINYPTL
 SALAPFLAERKGIQNGVEBPQAVESVSEBBDGAMLRLEOPQELSNEDLIGMLS
 DKXEDVSS"
 21463..26220
 /gene="sc1B"

Query Match 19.2% Score 817.2; DB 1; Length 66808;
 Best Local Similarity 57.5%; Pred. No. 2,4e-80;
 Matches 1597; Conservative 0; Mismatches 1148; Indels 31; Gaps 6;

QY 21 GCGCGCAGCCGAAGATCCGATTGCGATGTCGAGACGAGTTGCCCTGTCGCGGTGCGGT 80
 DB 43614 GCGTCTGGCCCCCGAGCCCATGCGATGCTGGGCACTGGCTGTCTGCTACCCCGTGGGGGT 43673
 QY 81 GATGATCTGAAGCGGGTCTTGACGCTCTTCGAGGGCTTCGCGCGACACCGTGGCGGAGT 140
 DB 43674 GAAGACATCCCACTCTATTGGGATTTGTTGAAGCGCGCGGAGATGCGGTTCACGGAAT 43733
 QY 141 CCGCGCGAAGCGTGGAGTGAAGACGCGTGTGATCCCAACCCCGATGCCCGCGGGAA 200
 DB 43734 TCCAGCAACCCGCGTGGAGATCAACGCTTCTTCATCCAGCAACGAGCAAGCCCGGCA 43793
 QY 201 GACGCCCTTGAAGCGCGATCTTTCTGAGCGACCTGACCTGCTTGAAGCTCTTCTT 260
 DB 43794 GACGACGAGCGGCTGGGGGGGCTTCTGATGATGATGAGGGGTTGATCCCAACCTGT 43853
 QY 261 CCGCATCTCGCTCGCGAAGCGCTGCGAGTGAACCTTCAACATGCACTCTTGCTGAGGT 320
 DB 43854 CCGCATCTCGCTCGCGAAGCGCGAGATGATCCCAACGCGCTCTTCTTGAGAGT 43913
 QY 321 GTGCTGGAGGCGCTGAGGAAGCGCGGATGCTCATCGGCGCTGCTGCTGATCGGAAC 380
 DB 43914 CGCCTGGAGGCGCTGAGGAAGCGCGAAGCAACGATGCTGTGCGCGGAGGAC 43973

QY 381 GGAAGTTCATCGGAGATCGGCGCCCTCCGAATATGAGCGCGCTGCCGCAAGCAGCGC 440
 DB 43974 GGGGGTCTTCGCGGCTTTTACAGAGGACTACTCTCTGTGTC-----TCTCGGA 44027
 QY 441 GTCCGCAAGATCCAGCTCATGCGCGGCTGGGACCATGCCACGCTTCGAGCGGCGCG 500
 DB 44028 TCTCAAGCCCATCAACGCTTATACCGGACAGGACCGGCAACACGCGGCGCTGAGCG 44087
 QY 501 AATCTCGATGCGCTCGGCGCTGCGAGGGCGGTGTGTCGCGGTGATACCGGCTATTTCGTC 560
 DB 44088 ATTGTCTATTTGCTGACCTTGCAAGGGCCAGACTTTGTGTGACACGGCTGCTCTTC 44117
 QY 561 CTCGCTGTGCGCGTTACCTGAGCTTGTGACAGGCTTGCGCTCCGCGAATCTCCACCGC 620
 DB 44148 CTCGCTGTGCGCGTTACCTGAGCTTGTGACAGGCTTGCGCTCCGCGAATCTCCACCGC 44207
 QY 621 CTCGCTGTGCGCGTTACCTGAGCTTGTGACAGGCTTGCGCTCCGCGAATCTCCACCGC 680
 DB 44208 GGTGCTGCGCGCGTTCAGCTTGTGCTCACCGCCACTGACCTTCATGAGCGCTGAGAT 44267
 QY 681 CCGGCGCGTGCAGAGGACGCTGCTGCAAGGCAATTTTCGCGGAGCGCGATGAGGTTGCG 740
 DB 44268 GCACTCATGCGCGCGAGCGCGCGCTGCAAGACGTTGACAGCGCGCGCTGATGCG 44327
 QY 741 ACGAGCGCAAGGTCGCGCGCTGCTGCTTCAAGCGGCTCAGTGAAGCCCGCGGACGG 800
 DB 44328 CATGGGTGAGGGCTGTGGCGGTGATGATCTCAAGCGGCTTTCGATGCGGACGGCGATG 44387
 QY 801 CGATTCGATATTGCGCGGTGATTCAGAGATCCGCGATCAATACAGCGGTGGAGCAGCGG 860
 DB 44388 TGAATCGATCTCTGCGCGGTGATCAAGGCTTCGCGCATCATACAGCGCGCGCAAGATG 44447
 QY 861 TCTGACGCTGCGCAAGCGGACGCTCCCAAGAAATGATGCTGAACCGGCGCTGCGGACGG 920
 DB 44448 GCTGACACGCGCGAGTGGCGCTTCCACAGAGGTGTTGTGCTCAGCGGCTGACCGCGCG 44507
 QY 921 AGGCTGCGCGCGTCTTCTGAGGTATGTCGAGGACACAGCGACGCGGACGAGCGCTTGG 980
 DB 44508 TGGCATCTCGCGGTGCGAGGTGAGGTAGCTGAGAGCCAGCGGTACGGGACACCGCTCGG 44567
 QY 981 TGACCCCATCGAAATCCAGCTTGAATGCGGTATAGCGCTTCGAGCGAGATGTGCGCAC 1040
 DB 44568 CGATCCGATGAGGTGAGGCGCTTCGCGAGGTATCGGAGCGACCAACGCTTCGCGGCG 44627
 QY 1041 GCGCGTGTGATTCGGGTGCGGAGGACCAACCTTGCGCATCTGAGTATGCGGTGGAGT 1100
 DB 44628 TCCCTGTGCTGAGGTTCGTAAGGCCAATCTGAGACATACCGCGCGCGCGGGGCT 44687
 QY 1101 CACTGCGTGTGTAAGGTGCTCTTGTCTCTTCAAGCAGCGGAGATTCCTGCGCACCTTCA 1160
 DB 44688 CCGCGGCGTATCAAAAGTCTGCTCTCATACAGAGGCGATTCCTCTCAACTGCA 44747
 QY 1161 CCGCAGGCGCTGAACCCCGGATCTCATGAGGTGATCTTGGCTGACCGTCAAGCGCGC 1220
 DB 44748 CTACCGGAGGCTCAACTCAATCTCTTGAAGAACCTCTCGTGTCAATTCGAGAGA 44807
 QY 1221 CCGGACACCGTGGCGGATGAGAAACGCGCGACGCGGCGGGGAGCTGTTTCGAGAT 1280
 DB 44808 GCTGATCTCTGCGCTGCGGAGAGGGCGCGGTATGACAGGATCACTGCTTCGTTG 44867
 QY 1281 GAGCGGACCAACCGCAGCTGATGCTGAGAGAGGCGCGGCGAGCTGACACACCGCC 1340
 DB 44868 GTACAGGAGCAACCGCAGCGTGTGAGAGAGGCGCGCTGCGCGCCCAAGGCTCGCGC 44927
 QY 1341 GCGCGCGAGCGACCGGCAAGCTGCTGTGCTGTGCGCAAGACCGGCTCAGCGCTTGA 1400
 DB 44928 GCGCGTGAAGGTCCGCGTGCACCTTCGACCTTGTGCGACCGAGGCGCGGCTGCG 44987
 QY 1401 TGCAAGGCGCGCGGCTGCGGACCAATCTGAGACCTTTCGAGTCTTGGGAGC 1460
 DB 44988 CGAGTTGCGGAGGCTGTGAAGACATATGTTTCGCGGACCGGAGATCTGCCATCGCGA 45047
 QY 1461 TGTGCGTTGATCTGCGAGACGCGCAGCGCATGAGAGCACCGGCTTCGCGTGGCGCGC 1520

Db	45048	TCTCTGCTACTCGGCGGAAC	CGGGGTGTGGAAGTTGAGGACCAATCTGGCGGTCCCGC	45107
Oy	1521	GACGTCGAGGAGGGGCTGCGGCGAC	CCCTTGACGCTGCGGCGCAGAGGACAGACGTGCGC	1580
Db	45108	GGGGAACAAGGCGCAGATTGAGCGAG	CGTCTCTCCGCGTTTCTTGAGGGGACAGCCGACAGC	45167
Oy	1581	CGGTGCGGTGCGGAGTATGCGCCGAT	TTCCTCAACGCGGCAAG---CTGGCCTTTCCTTCAAC	1637
Db	45168	AGGCACTCTGTAAGGACAGGTCA	CGGGACGCGCGTGGAGCGGGAGTGGCGTTCTGTTCAC	45227
Oy	1638	CGGACAGGGGCGCAGACGCTGGG	CATGGGCGCTGGAGCTGTACATGTATGTCGCGCTT	1697
Db	45228	GGGGCAAGGCGCTCGAGTACCGGG	CAATGGGGCGGCACTTACCAACACAGCCCAATT	45287
Oy	1698	CCGGGAGCGCTTCGACCTGTGCG	TGGAGGCTTTCAACAGAGAGCTCGACCGGCGCTCCG	1757
Db	45288	CCGGCGTGCCTCGATCGGTGCGA	CGAGTTGTGCGGCCCTTGTCTCGGGCGCTCAATTCT	45347
Oy	1758	CGAGGTGATGTGGGCGGACCGG	CGACGCTCGACGCGCTCTCGACACAGACACTT	1817
Db	45348	GTCTGTATCTACCCCAAGG----	GGGACAGGCTTCTCCATGACAGAGCGGCTT	45401
Oy	1818	CACCCAGCGGCGCTGTTCACCTT	CGAATATGCGCTCGCGCGCTGTGGCGGCGTGGGG	1877
Db	45402	CACCCAGCGCGCTTTGTTCGCTT	CGAGTGGGCGCTCGGAACTCTGGCGCTGTGGGG	45461
Oy	1878	TGTAGACGCGGAGTTGTGTGCG	CGGCGCAATACATCGGTAGACTGTGGCTGTGCG	1937
Db	45462	CGTTGTACCGACCGTGTGATGG	GGCCACAGCGTGGGTATGTGGCGGCGTGGCTGCG	45521
Oy	1938	GGGCGGTGTTCTCGCTTGAAGA	ACGCGGTGTTCTGTGTGCTCGCGCGGGCGCTGATGCA	1997
Db	45522	TGGGTGTTCACCTTCAGAAAGG	CGTGGCGTCACTGTGTGAACGCGCGCGGTGTATGCA	45581
Oy	1998	GGCGCTGCGGCGCGGGCGAGT	GTGTGATCGAGGCGCGGAGGCGGATGTGGCTGC	2057
Db	45862	GTCTGTGCGCGGGCGGCGAGAT	GGCGGCTTTCACGAGCGGGCTGTGACCGCGCA	45641
Oy	2058	TGCGGTGGCGCGCAGCAGCGT	CGGTGATCGCGCGCGGTCAAGCTCCGAGCAGGT	2117
Db	45642	GGCGATTGTCTTAAACGATCG	ACAGGTCTCATTTGCGGCTTCAATGTGCTTTCCGAGAC	45701
Oy	2118	GGTCAATCGCGGGCGCGGCAAC	CGGTGCATTCGACGCGGCGCAATGGCGCGCGCGG	2177
Db	45702	GGTATCTCTTGGCATGGAAGG	CGGTGGAGCGGATTTCTGGGGCGTTTCCGCCAGAGG	45761
Oy	2178	GGCGCGAACCAAGCGCGTCA	CGTCTGCGCATGCGTCACTACCGCTCAATGGCGCGCAT	2237
Db	45762	CTGTAAAGTTTCGGGCGCTGG	GTCTTCCACAGCTTTCACGTCTTCGCTCATGTGACCCAT	45821
Oy	2238	GCTGAGGCGTTTCGGGCGTGT	GGCGAGTGTGCGGTAGCTACCGCGGCGCGTCACTCGTCT	2297
Db	45822	GCTGGATGCTTTCAGCGGGG	CGGGGAGAACGCGCTTCAATCCGCCAAGATCACGT	45881
Oy	2298	GGTACCAATCTGAGCGGAAAG	GGCTTGACACAGACGAGGTGAGCTGCCCGGGCTAATTGGT	2357
Db	45882	GATCTCGAACCTACCGG-----	TGGGCGCTCGAGAGAGTTCTCGGCGCGGGCGTTTGGG	45931
Oy	2358	GGCGCAACGCGGAGAGGTGTG	CGCTTCCGGATGTGAGTGAAGGCCCTGCAACGCGGCGG	2417
Db	45936	TCGACATGCGCGTAAACGG	GTCCGTTCTCGAATGTGGCGGCTCAAGGCAAGTGG	45991
Oy	2418	TGCGGAGCACTTTCGTGAGG	TCGGTTCGGAATGACGCTGTGCGGCGTGGTGCCTGCTG	2477
Db	45996	CGTCTCGGTGCGGTGAGATT	TGGCGCCCAATGTCACTGTCTCGGATCGGAAGCGGATG	46055
Oy	2478	CATGCGGAGCGCCCGCGCGG	CGCTGTGATCGTGTGCGCGCTGTGGCGTGAACGAGCGGCG	2537
Db	46056	TCTGCGGAAAGGTTCGGGT	GTGTCCTTTCGAAAGGACAAAGACGAGTGGGA	46115
Oy	2538	GACCGTCTCGAGGCGCTCG	CGGCGCTCTGGCGGTGGCGCTGGTCTCTGGGCGG	2597

Db	46116	AGTCTGCTCGCAACGTTGGGTGCCTCTGTGNCAGGACGTTCCGTCACAATCGGACAG	46117
OY	2598	-----CCTTTCCTCCCTCAGAGGGGGGGGGGGTCCGCTGCGGCCACTACTCTTGGAAGGGCA	2654
Db	46176	GTTTCGACAGGACACTAACCCGGTGGGGGGGTCTCCGCCCCAATTATTCGTTTCAGACGA	46235
OY	2655	GCGCTACTGTGATGCACNCAAACCAGCACGCGGCGCGTGGCGACCGCGCTGCTCCGG	2714
Db	46236	GCGGTTTGCCTC-----AAACCAATGATGATGAGAGGGGTGTGGGCGCGAGCGGTCAAG	46288
OY	2715	AGCGGGTTCACAGACGAGTTCGAGGAGGGGGCGCGGTGCGCGCGGCGACCGCGCAGCGC	2774
Db	46289	TGCGGTGTCATCCCTGTGGATTCGGCTGTGCGCTGCGCGTCCGACGCGGTGACGGCATCTGT	46348
OY	2775	TGCGCTCGACCATTCG 2790	
Db	46349	TGCAATCGAAGTTCCG 46364	
RESULT 15			
AR266876		3978 bp	DNA linear PAT 10-APR-2003
LOCUS	AR266876		
DEFINITION	Sequence 19' from Patent US 6495348.		
ACCESSION	AR266876		
VERSION	AR266876.1 GI:29696258		
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 3978)		
TITLE	Sherman,D.H., Mao,Y., Varoglu,M., He,M. and Sheldon,P.		
JOURNAL	Mitomycin biosynthetic gene cluster		
FEATURES	Patent: US 6495348-A 19 17-DEC-2002;		
source	Location/Qualifiers		
	1..3978		
	/organism="unknown"		
BASE COUNT	497 a 1583 c 1415 g . 483 t		
ORIGIN			
Query Match	19.1%; Score 814.2; DB 6; Length 3978;		
Best Local Similarity	57.6%; Pred. No.1,2e-79;		
Matches 1549; Conservative	0; Mismatches 1108; Indels 30; Gaps 4;		
OY	18	CGAGCGCGCGACCGCAAGATCCGATTCGATCTGCGAGCGAGTGGCGCTGCGCGGTGG 77	
Db	90	CGAGGAGCGCGGCCCGCGACGCCGTGCGATCTGGGCGATGGCGTCCCTTCCCGGGGA 149	
OY	78	CGTGATGATCTGAGCGGGTTCTGAGCGCTCTCGAGGGCTGCGCGACACGTCGGGCG 137	
Db	150	CGTGGCATTCGCGGACCACTGTGGCAGCTGTGTCGCCGAGGGCGGGAGCGCCTTACCGA 209	
OY	138	AGTCCCGCGCGAAC--GCTGGAGATGCACGAGCGTGTGATCCCGACCCCGATGCCCC 194	
Db	210	GTTCCCGCGGACCGGGGGCTGGGACGTGACGCGCGTACTAGACCCCGAGCGGGGACCCC 269	
OY	195	GGGGAAGACCCCGGTTACGCGCGCATTTTCTTGAGCAGATGAGCTGCTTGAAGCGCTTC 254	
Db	270	GGGACAGACGTAACGCGCGCACCGCGGCTTCTTCAAGAGACGCGCGCGATTGCAAGCCCG 329	
OY	255	CTTCTTGGGATCTCGGCTGCGGAGAAGCGCTGCGGATGAGACCCCGACATTCGACTTGTCT 314	
Db	330	CTTCTTGGGATCAACGCGCGGAGGGGCTGCGCATGAGACCGCAGCAGGCAATGATCAT 389	
OY	315	GGAGGTGCTGGAGGCGCTGAGAGAACCGCGATTCGCTCATTCGGCGCTCGTCGTAC 374	
Db	390	GGAGGTCTCTCGGAGGCGTTTGAAGAGGGGGCGCTTGAACGACCAACCTTGGGGGCGA 449	
OY	375	GGAAACGGAGATGTTCAATCGGAGATCGGCCGCTCGCAATATGAGAGCGCGCTGCGCAACG 434	
Db	450	GGAAGTGGGGGCTTTCGTGGCTCCAACACCAACGACTACTGATCAACGTGCTCGACGC 509	
OY	435	GACGGCGTCCGACAGATCAACGCTATGGCGGGGCTGGGAGCAATCCAGCCTCGGAGC 494	

Dh 510 GCGGAGCGTGCAGAGGCTTC-----ATCGGAGACCGGCAACTCCCGACGATCTCTC 563
Qy 495 GGGCCGATCTCGATGCGCTCGGGCTTCGAGGGCCGCTGCTCGGGGATACGAGCTA 554
Dh 564 CGGCGCGTGCCTTACACCTTCGAGCTTCGAGGGCCCGGCGCTGCTCGGACACCGCTG 623
Qy 555 TTGCTCTCGTGTGCGCTTCACTGAGCTTCGAGCTTCGAGCTTCGAGCTTCGAGCTTC 614
Dh 624 CTCTCTCTCGTGTGCGCTTCGAGCTTCGAGCTTCGAGCTTCGAGCTTCGAGCTTCGAG 683
Qy 615 CAGCGCTCTGCTGTGAGGATGCTGATGCTGTGCTGAGGACCTCTGCTGCTGCTC 674
Dh 684 CTGAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 743
Qy 675 GAGAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 734
Dh 744 CGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 803
Qy 735 GTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 794
Dh 804 CACCACTGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 863
Qy 795 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 854
Dh 864 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 923
Qy 855 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 914
Dh 924 CGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 983
Qy 915 GAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 974
Dh 984 CAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1043
Qy 975 GCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1034
Dh 1044 GCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1103
Qy 1035 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1094
Dh 1104 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1163
Qy 1095 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1154
Dh 1164 CGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1223
Qy 1155 CTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1214
Dh 1224 GCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1283
Qy 1215 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1274
Dh 1284 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1343
Qy 1275 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1334
Dh 1344 CGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1403
Qy 1335 ACCGCG-----GCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1376
Dh 1404 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1463
Qy 1377 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1436
Dh 1464 CGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1523
Qy 1437 CTACCTTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1496
Dh 1524 GCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1583
Qy 1497 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1556
Dh 1584 CGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1643

Qy 1557 TCGGCGCAGAGGAGCAGAGCTGCGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1616
Dh 1644 CTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1703
Qy 1617 CAGGCTGCGCTTCTTTCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1676
Dh 1704 CAGGCGGCTTCTTTCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1763
Qy 1677 GTACGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1726
Dh 1764 GCGGCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1823
Qy 1737 GAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1796
Dh 1824 GAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1883
Qy 1797 GCTGCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1856
Dh 1884 CTGCTGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1943
Qy 1857 CGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1916
Dh 1944 GCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2003
Qy 1917 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1976
Dh 2004 ACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2063
Qy 1977 TCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2036
Dh 2064 CGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2123
Qy 2037 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2096
Dh 2124 GAG 2183
Qy 2097 GATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2156
Dh 2184 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2243
Qy 2157 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2216
Dh 2244 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2303
Qy 2217 CTGAG 2276
Dh 2304 CTGAG 2363
Qy 2277 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2336
Dh 2364 CAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2423
Qy 2337 GAGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2396
Dh 2424 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2483
Qy 2397 GAG 2456
Dh 2484 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2543
Qy 2457 GCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2513
Dh 2544 GAG 2603
Qy 2514 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2573
Dh 2604 GCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2663
Qy 2574 CGGAG 2633
Dh 2664 CGGAG 2723

Qy 2634 CACGTACCCCTTGACAGCGGAGCGCTACTGATCGACAGAAAGCCG 2680
Db 2724 CACGTACCGCTCTTCACAGCGCGCCACTACTGCTGCGGCGATGACCg 2770

Search completed: October 4, 2003, 03:30:21
Job time : 15291.8 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 2, 2003, 17:27:29 ; Search time 125.882 Seconds
(without alignments)
2912.986 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210
Sequence: 1 VADPPIERADPIAIVGAS.....GVNDPVSSGADQDWETIAL 1421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7207	100.0	1421	2	Q918C9
2	7081	98.2	1421	2	Q9KJ00
3	4137	57.4	2439	2	Q918C5
4	4098	56.8	2439	2	Q9K1Z5
5	3322.5	46.1	7257	2	Q9K1Z7
6	3311.5	45.9	7257	2	Q918C7
7	2634	36.5	2218	2	Q8BJY1
8	2606	36.1	6315	2	Q9ADL6
9	2416.5	33.5	5017	2	Q8GBX6
10	2357	33.7	5644	2	Q93NX8
11	2355.5	33.7	5435	2	Q93NX2
12	2331.5	33.3	11096	2	Q914W3
13	2330	33.3	9507	2	Q9EWA1
14	2328.5	32.3	10917	2	Q93NW6
15	2321	32.2	4151	16	Q53450
16	2313	32.1	3513	2	Q8GBX5

17	2311.5	32.1	6146	2	Q93HJ5	Q93HJ5 streptomyc
18	2298.5	31.9	1203	16	Q9CD81	Q9CD81 mycobacteri
19	2275.5	31.6	2159	2	Q8RJX8	Q8RJX8 stigmatella
20	2273	31.5	2152	2	Q9ALM5	Q9ALM5 saccharopol
21	2262.5	31.4	1885	2	Q8RJY4	Q8RJY4 stigmatella
22	2260.5	31.4	9477	2	Q914X3	Q914X3 streptomyc
23	2260	31.3	8817	2	Q93840	Q93840 polyanthum
24	2258	31.3	9510	2	Q93NX9	Q93NX9 streptomyc
25	2257	31.3	1604	2	Q8RJY9	Q8RJY9 stigmatella
26	2255.5	31.3	1585	2	Q8RJY5	Q8RJY5 stigmatella
27	2249	31.2	6145	2	Q93H84	Q93H84 streptomyc
28	2245.5	31.1	3798	2	Q918C6	Q918C6 polyanthum
29	2238.5	31.0	3798	2	Q9K1Z6	Q9K1Z6 polyanthum
30	2229.5	30.9	5192	2	Q93TW9	Q93TW9 stigmatella
31	2211.5	30.7	2478	16	Q8JTN5	Q8JTN5 anabaena sp
32	2207	30.6	1602	16	Q65933	Q65933 mycobacteri
33	2199.5	30.5	3352	2	Q93H83	Q93H83 streptomyc
34	2174	30.2	3579	2	Q8GBX4	Q8GBX4 polyanthum
35	2165	30.0	6797	2	Q9X993	Q9X993 streptomyc
36	2157.5	29.9	1360	2	Q9RFK6	Q9RFK6 stigmatella
37	2157.5	29.9	3524	2	Q93H86	Q93H86 streptomyc
38	2156.5	29.9	6048	2	Q93H87	Q93H87 streptomyc
39	2152	29.8	1835	2	Q93H85	Q93H85 streptomyc
40	2150	29.8	1841	2	Q93958	Q93958 streptomyc
41	2136	29.6	2518	16	Q8JTN4	Q8JTN4 anabaena sp
42	2128	29.5	3816	2	Q9K1V3	Q9K1V3 streptomyc
43	2127	29.5	3170	2	Q9ALM4	Q9ALM4 saccharopol
44	2123.5	29.5	2024	2	Q9EWA3	Q9EWA3 streptomyc
45	2120	29.4	1587	16	Q8JW65	Q8JW65 anabaena sp

ALIGNMENTS

RESULT 1

ID	Q918C9	PRELIMINARY;	PRT;	1421 AA.
AC	Q918C9;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Polyketide synthase.			
GN	EPOL.			
OS	Polyanthum cellulorum.			
OC	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;			
OC	Sorangineae; Polyanthaceae; Polyanthum.			
OX	NCBI_taxid=56;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=So ce90;			
RX	MEDLINE=20130945; PubMed=10662695;			
RA	Molnar I., Schupp T., Ono M., Zirkle R.E., Milnaw M.,			
RA	Nowak-Thompson B., Engel N., Toupet C., Stratzmann A., Cyr D.D.,			
RA	Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;			
RT	"The biosynthetic gene cluster for the microtubule-stabilizing agents			
RT	epothilones A and B from Sorangium cellulorum So ce90."			
RL	Chem. Biol. 7:97-109 (2000).			
DR	EMBL; AF210843; AAF26919.1; -			
DR	HSSP; P25715; IMLA.			
DR	InterPro; IPR001227; AC transferase.			
DR	InterPro; IPR002085; Adh_zn family.			
DR	InterPro; IPR000794; Ketoacyl-synt.			
DR	InterPro; IPR006162; Pantone_attech.			
DR	InterPro; IPR006163; PP_bind.			
DR	Pfam; PF00698; Acyl_transf.1.			
DR	Pfam; PF00107; adh_zinc.1.			
DR	Pfam; PF00109; ketoacyl-synt.1.			
DR	Pfam; PF02801; ketoacyl-synt.C.1.			
DR	Pfam; PF00550; pp-binding.1.			
DR	PROSITE; PS50075; ACP DOMAIN.1.			
DR	PROSITE; PS00012; PROSPHANTHETINE.1.			
DR	Phosphatetheine.			
KW	SEQUENCE 1421 AA; 149035 MW; ABB5615EDAEDC996 CRC64;			

Query Match 100.0%; Score 7207; DB 2; Length 1421;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1420; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRPETRAEDPIALVAGASCRLPGVVIDISGFMTLLBESRDYVGRVPERMDAAMPDP 60
 1 MADRPETRAEDPIALVAGASCRLPGVVIDISGFMTLLBESRDYVGRVPERMDAAMPDP 60
 DB 61 DDPAPKPTVTRASFISDVACPDASFPGISPREALMDPAHLLLEVCHEALENAAIAPS 120
 61 DDPAPKPTVTRASFISDVACPDASFPGISPREALMDPAHLLLEVCHEALENAAIAPS 120
 QY 121 ALVGTETGVFTIGSPSEYEALPQATASAEIDAHGGIGTMSVAGRISVALGRPCVA 180
 121 ALVGTETGVFTIGSPSEYEALPQATASAEIDAHGGIGTMSVAGRISVALGRPCVA 180
 DB 121 ALVGTETGVFTIGSPSEYEALPQATASAEIDAHGGIGTMSVAGRISVALGRPCVA 180
 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMSPSTLWLSKTRALARDGRCKAS 240
 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMSPSTLWLSKTRALARDGRCKAS 240
 QY 241 AEADEGFRGEGCAVVLKRLSGARADGRIILAVIRGSAINHGCASSGLTPNGSSQEIYL 300
 241 AEADEGFRGEGCAVVLKRLSGARADGRIILAVIRGSAINHGCASSGLTPNGSSQEIYL 300
 DB 241 AEADEGFRGEGCAVVLKRLSGARADGRIILAVIRGSAINHGCASSGLTPNGSSQEIYL 300
 301 KRALLADGCAASSVGVVEAHGTGTLGDPTEIQALNNAVGLGRDVATPILLISGVKTNLGH 360
 301 KRALLADGCAASSVGVVEAHGTGTLGDPTEIQALNNAVGLGRDVATPILLISGVKTNLGH 360
 QY 361 PEVASGITGLLKVLSLQHQIIPAHLLAQLNPRISWGDRLTVTRARPTWPMNTPRRA 420
 361 PEVASGITGLLKVLSLQHQIIPAHLLAQLNPRISWGDRLTVTRARPTWPMNTPRRA 420
 DB 361 PEVASGITGLLKVLSLQHQIIPAHLLAQLNPRISWGDRLTVTRARPTWPMNTPRRA 420
 421 GVSSFGMSGTNAVLTBEAPAACTTPAPERPALLVLSAKRTASALDAQARLDHLETY 480
 421 GVSSFGMSGTNAVLTBEAPAACTTPAPERPALLVLSAKRTASALDAQARLDHLETY 480
 QY 481 PSCGLDVAFLSLTTSAMEHRLAVATSRGRLAALDAAOQOTSFGAARSLADSSRG 540
 481 PSCGLDVAFLSLTTSAMEHRLAVATSRGRLAALDAAOQOTSFGAARSLADSSRG 540
 DB 481 PSCGLDVAFLSLTTSAMEHRLAVATSRGRLAALDAAOQOTSFGAARSLADSSRG 540
 541 LAFLFTGQGAQTLGMRGLYDVSAFREAFLDLCVRLFNQELDRPLREVMMAEBSYDAL 600
 541 LAFLFTGQGAQTLGMRGLYDVSAFREAFLDLCVRLFNQELDRPLREVMMAEBSYDAL 600
 QY 601 LDDTAFTQPALFTFEVYLAALMRSWGVPELVAGHSGELVAAACVAGVSELEDAVFLVAA 660
 601 LDDTAFTQPALFTFEVYLAALMRSWGVPELVAGHSGELVAAACVAGVSELEDAVFLVAA 660
 DB 601 LDDTAFTQPALFTFEVYLAALMRSWGVPELVAGHSGELVAAACVAGVSELEDAVFLVAA 660
 661 RGRIMQALPAGGAMVSTIEAPEADVAAPAAVAPHAASVSTIAANAPDQVVIAGAGOPVHAIAA 720
 661 RGRIMQALPAGGAMVSTIEAPEADVAAPAAVAPHAASVSTIAANAPDQVVIAGAGOPVHAIAA 720
 QY 721 AMAARGARTALVSHAHSPLMAFMLAEFVAESVSYRPSIVLVSNLSGKACTDEVS 780
 721 AMAARGARTALVSHAHSPLMAFMLAEFVAESVSYRPSIVLVSNLSGKACTDEVS 780
 DB 721 AMAARGARTALVSHAHSPLMAFMLAEFVAESVSYRPSIVLVSNLSGKACTDEVS 780
 781 SPGYWRHAAREVVRPADGVATLHAAGAGTVEVGPSTLLGLVPAQCPDARPALASSRA 840
 781 SPGYWRHAAREVVRPADGVATLHAAGAGTVEVGPSTLLGLVPAQCPDARPALASSRA 840
 QY 841 GRDEPRTVLALGLVAVGLVSWAGLPSPGGRVPLPTYPKMRERWIDTKADDARGD 900
 841 GRDEPRTVLALGLVAVGLVSWAGLPSPGGRVPLPTYPKMRERWIDTKADDARGD 900
 DB 841 GRDEPRTVLALGLVAVGLVSWAGLPSPGGRVPLPTYPKMRERWIDTKADDARGD 900
 901 RRAFGAGHDEVEEGAVRGDRSARLDHPPSPSGRREKTEAAGDRPFRLLEIDEPVLDH 960
 901 RRAFGAGHDEVEEGAVRGDRSARLDHPPSPSGRREKTEAAGDRPFRLLEIDEPVLDH 960
 QY 961 LVRVTRERRAPGIGVEIIVADAAGLSNDVOLALGVVDDLPCKGNPPLLIGECACRIV 1020
 961 LVRVTRERRAPGIGVEIIVADAAGLSNDVOLALGVVDDLPCKGNPPLLIGECACRIV 1020
 DB 961 LVRVTRERRAPGIGVEIIVADAAGLSNDVOLALGVVDDLPCKGNPPLLIGECACRIV 1020

QY 1021 AVEGCVNGLVGQPVIALSAGAFATHVTTSALVLPPOQASIAEAAAMPVATLAWVAL 1080
 1021 AVEGCVNGLVGQPVIALSAGAFATHVTTSALVLPPOQASIAEAAAMPVATLAWVAL 1080
 DB 1081 DRIARLOPGERVLIHAATCGVGLAAVQMAQHVGAEVHATAGTEPKRAYLESIGVRYVSDS 1140
 1081 DRIARLOPGERVLIHAATCGVGLAAVQMAQHVGAEVHATAGTEPKRAYLESIGVRYVSDS 1140
 QY 1141 RSDRFVADVARTAGGEGVDVNVNSLSGELIDSFNLLSHGRFVELGKRDCAVDNQLGR 1200
 1141 RSDRFVADVARTAGGEGVDVNVNSLSGELIDSFNLLSHGRFVELGKRDCAVDNQLGR 1200
 DB 1141 RSDRFVADVARTAGGEGVDVNVNSLSGELIDSFNLLSHGRFVELGKRDCAVDNQLGR 1200
 1201 PFLRNLSFSLVDLRGMLEPRPARVRLAEELGLIAAGVTPPPIATLPIAVADAFRSM 1260
 1201 PFLRNLSFSLVDLRGMLEPRPARVRLAEELGLIAAGVTPPPIATLPIAVADAFRSM 1260
 QY 1261 AQAQHLKGLVLTGDPVEOIRIPTIAGACPSGTGRDLDRLASAAPARAALAEFLRQ 1320
 1261 AQAQHLKGLVLTGDPVEOIRIPTIAGACPSGTGRDLDRLASAAPARAALAEFLRQ 1320
 DB 1261 AQAQHLKGLVLTGDPVEOIRIPTIAGACPSGTGRDLDRLASAAPARAALAEFLRQ 1320
 1321 VSOVARTPEIKVGAELFTRLGMDSLMAVELRNRIEASIKLSTFTSTSPNIALAON 1380
 1321 VSOVARTPEIKVGAELFTRLGMDSLMAVELRNRIEASIKLSTFTSTSPNIALAON 1380
 QY 1381 LLDALATALSLEVAENIRAGVONDPVSSGADQDMEITAL 1421
 1381 LLDALATALSLEVAENIRAGVONDPVSSGADQDMEITAL 1421
 DB 1381 LLDALATALSLEVAENIRAGVONDPVSSGADQDMEITAL 1421

RESULT 2
 09KJ00 PRELIMINARY; PRT; 1421 AA.
 ID 09KJ00;
 AC 09KJ00;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE EgoA.
 GN EPOA.
 OS Polyangium cellulosum.
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
 OC Sorangineae; Polyangiaceae; Polyangium.
 OX NCBI_TaxID=56;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SMP44;
 RX MEDLINE=20283058; PubMed=10831849;
 RA Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
 RT "Isolation and characterization of the epoIII gene cluster from Sorangium cellulosum.";
 RL Gene 249:153-160(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SMP44;
 RX MEDLINE=20115953; PubMed=10649995;
 RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C.;
 RT "Cloning and heterologous expression of the epoIII gene cluster.";
 RL Science 287:640-642(2000).
 DR EMBL; AF217189; AAF62880.1; -.
 DR HSSP; P25715; IMLA.
 DR InterPro; IPR001227; AC_transferase.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR006162; Pantane_attach.
 DR InterPro; IPR006163; PP_bind.
 DR Pfam; PF00698; Acyl_transf.1.
 DR Pfam; PF00107; adh_zinc.1.
 DR Pfam; PF00109; ketoacyl-synt.1.
 DR Pfam; PF02801; ketoacyl-synt.C.1.
 DR Pfam; PF00500; pp-binding.1.
 DR PROSITE; PS50075; ACP_DOMAIN.1.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE.1.
 KW Phosphopantetheine.
 SQ SEQUENCE 1421 AA; 148744 MW; 3D055DDBF686E2C5 CRC64;

Query Match 98.2%; Score 7081; DB 2; Length 1421;
 Best Local Similarity 98.2%; Pred. No. 0;
 Matches 1396; Conservative 7; Mismatches 18; Indels 0; Gaps 0;

1 VADRPERRAEDPIAIVGASCRLEPGVINDLSGFVTLLEGSRDTGVGVPERMDAAEDP 60
 1 MADRPERRAEDPIAIVGASCRLEPGVINDLSGFVTLLEGSRDTGVGVPERMDAAEDP 60
 61 DDAPKPTVTBASFLSDVACPDSPFGISPREALMDPAHRLLEVCHEALENAAIAS 120
 61 DDAPKPTVTBASFLSDVACPDSPFGISPREALMDPAHRLLEVCHEALENAAIAS 120
 121 ALVGETGVFIGIGPEYEVALPOATASAEIDAHGGLGTMPSVAGARISYALGRPCVA 180
 121 ALVGETGVFIGIGPEYEVALPOATASAEIDAHGGLGTMPSVAGARISYALGRPCVA 180
 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLVLSKTRALATDGRKAPS 240
 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLVLSKTRALATDGRKAPS 240
 241 AADGGRGEGCAVVVLKRLSGARADGRILAVIRGSAINHOGASSGLTVPNGSSQEV 300
 241 AADGGRGEGCAVVVLKRLSGARADGRILAVIRGSAINHOGASSGLTVPNGSSQEV 300
 301 KEALADAGCAASSVGYEAHGTGTLGDPRIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
 301 KEALADAGCAASSVGYEAHGTGTLGDPRIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
 361 PEVASGITLLKLVLSLQHGQIPAHILHAQALNPRISMGDLRLTVTRARTPMDMTPRRA 420
 361 PEVASGITLLKLVLSLQHGQIPAHILHAQALNPRISMGDLRLTVTRARTPMDMTPRRA 420
 421 GVSSRGMSSTNAHVLEEAIPAATCTPAPERPPELLVLSARTASALDAQAARLDLETRY 480
 421 GVSSRGMSSTNAHVLEEAIPAATCTPAPERPPELLVLSARTASALDAQAARLDLETRY 480
 481 PSQCLGDVAFSLATTSAMEHRLAVAATREGLRALDAQAQOTSPGAVRSIADSSRCK 540
 481 PSQCLGDVAFSLATTSAMEHRLAVAATREGLRALDAQAQOTSPGAVRSIADSSRCK 540
 541 LAFLFTGGAGQTLGMRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMABEASVDAAL 600
 541 LAFLFTGGAGQTLGMRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMABEASVDAAL 600
 601 LDQTAFTOPALFTFEVALALMWSMGVEBELVAGHSIGELVAACVAGVSLDAVFLVAA 660
 601 LDQTAFTOPALFTFEVALALMWSMGVEBELVAGHSIGELVAACVAGVSLDAVFLVAA 660
 661 RGRLOALPAGGAMVSI EAPEADVAAVAAPHAASVSI AAVNAPDQVVIAGAGOPVAAIA 720
 661 RGRLOALPAGGAMVSI EAPEADVAAVAAPHAASVSI AAVNAPDQVVIAGAGOPVAAIA 720
 721 AMAARGARTKALHVSHPSPMLAEMLEAFGRVAESVYRPSIVLVSMLSGKACTDEVS 780
 721 AMAARGARTKALHVSHPSPMLAEMLEAFGRVAESVYRPSIVLVSMLSGKACTDEVS 780
 781 SGGYWRHAREVVRFPADGVKALHAAGAGTFVEVGPSTLLGLVPACMPDARPLLASSSA 840
 781 SGGYWRHAREVVRFPADGVKALHAAGAGTFVEVGPSTLLGLVPACMPDARPLLASSSA 840
 841 GRDEPATVLEALGGLVAVGLVSMAGLFRSGGRVPLPTYPMOKREYVLDTRADDAARD 900
 841 GRDEPATVLEALGGLVAVGLVSMAGLFRSGGRVPLPTYPMOKREYVLDTRADDAARD 900
 901 RBAPEGAGHDEVEEGAVRGDRRSARLDHPPEPSGRREKVEAAGDRPFLAIDEPCVLDH 960
 901 RBAPEGAGHDEVEEGAVRGDRRSARLDHPPEPSGRREKVEAAGDRPFLAIDEPCVLDH 960
 961 LVLRVTERAPGLGVEIIVADAAGLSFNVOQLAGVPPDLPRKPNPPLLIGECGGRIV 1020
 961 LVLRVTERAPGLGVEIIVADAAGLSFNVOQLAGVPPDLPRKPNPPLLIGECGGRIV 1020

1021 AVEGNGVLVGOPIYALSAGAFATHVTSALVLRPOALSIAEAAAMPVAYLTAWVAL 1080
 1021 AVEGNGVLVGOPIYALSAGAFATHVTSALVLRPOALSIAEAAAMPVAYLTAWVAL 1080
 1081 DRIALQGERVLIHAATGCGVGLAAVQNAOHVGAHVATAGTPERKAYLESIGVRYVSS 1140
 1081 DRIALQGERVLIHAATGCGVGLAAVQNAOHVGAHVATAGTPERKAYLESIGVRYVSS 1140
 1141 RSDRFVADVRAWTGEGVDVUNSLSGELIDSFWLSSHGRFVEIGKRDCAVDNOLGR 1200
 1141 RSDRFVADVRAWTGEGVDVUNSLSGELIDSFWLSSHGRFVEIGKRDCAVDNOLGR 1200
 1201 PFLRNLSFSLVDLRGMLEERPARVALLEBLGLIAAGVTPPPIATLPIARVADAFRM 1260
 1201 PFLRNLSFSLVDLRGMLEERPARVALLEBLGLIAAGVTPPPIATLPIARVADAFRM 1260
 1261 AQAQHLGKLVLTGPEVQIRPTAGAGPSGDRDLDRLASAAPAAARAALEAFLRQ 1320
 1261 AQAQHLGKLVLTGPEVQIRPTAGAGPSGDRDLDRLASAAPAAARAALEAFLRQ 1320
 1321 VSQVLRTEPKVGAELFTRLGMDSLMAVELNRLEASIKLSTTFSTSPNIALTON 1380
 1321 VSQVLRTEPKVGAELFTRLGMDSLMAVELNRLEASIKLSTTFSTSPNIALTON 1380
 1381 LLDALATALSLERVAENLRAGVQNDPVSSGADQDMEIIAL 1421
 1381 LLDALATALSLERVAENLRAGVQNDPVSSGADQDMEIIAL 1421

RESULT 3
 ID 0918CS PRELIMINARY; PRT; 2439 AA.
 AC 0918CS;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Polypeptide synthase.
 OS EPOE.
 OS Polysaccharide synthase.
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
 OC Sorangium; Sorangium; Sorangium; Sorangium; Sorangium;
 OC NCBI_Taxid=56;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=So ce90;
 RX MEDLINE=20130945; PubMed=10662695;
 RA Molnar I., Schnupp T., Ono M., Zirkle R.E., Milamow M.,
 RA Nowak-Thompson B., Engel N., Toupet C., Straumann A., Cyr D.D.,
 RA Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
 RT "The biosynthetic gene cluster for the microtubule-stabilizing agents
 RT epothilones A and B from Sorangium cellulosum So ce90.";
 RL Chem. Biol. 7:97-109(2000).
 DR EMBL; AF210843; AAF26923.1; -.
 DR HSP; P25715; IMLA.
 DR InterPro; IPR001227; AC transferase.
 DR InterPro; IPR002085; Adh. zn. family.
 DR InterPro; IPR00794; Ketocacyl-synt.
 DR InterPro; IPR006162; Pentate_attach.
 DR InterPro; IPR00163; Pp_bind.
 DR InterPro; IPR001031; Thioesterase.
 DR Pfam; PF00698; Acyl_transf. 1.
 DR Pfam; PF00107; adh_zinc. 1.
 DR Pfam; PF00109; ketoacyl-synt. 1.
 DR Pfam; PF02801; ketoacyl-synt_C. 1.
 DR Pfam; PF00550; pp-binding. 1.
 DR Pfam; PF00975; Thioesterase. 1.
 DR PROSITE; PS50075; ACP_DOMAIN. 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE. 1.
 DR PROSITE; PS00012; PHOSPHOPANTHERINE. 1.
 KM Phosphopantetheine; Transferase.
 SQ SEQUENCE 2439 AA; 257732 MW; 08A32D75BCBAD72C CRC64;

Query Match

57.4%; Score 4137; DB 2; Length 2439;

RESULT 4
ID 09K125 PRELIMINARY; PRT; 2439 AA.
AC 09K125;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE EPOF.
GN EPOF.
OS Polyangium cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangineae; Polyangiaceae; Polyangium.
OX NCBI_TaxId=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RX MEDLINE=20293058; PubMed=10831849;
RA Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT "Isolation and characterization of the epoH gene cluster from Sorangium cellulosum.";
RL Gene 249:153-160(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RX MEDLINE=20115953; PubMed=10649995;
RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT "Cloning and heterologous expression of the epoH gene cluster.";
RL Science 287:640-642(2000).
DR BMBL; AF217189; AAF62885.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR006162; Ppantn_attach.
DR InterPro; IPR001031; Thioesterase.
DR Pfam; PF006598; Acyl_cranef. 1.
DR Pfam; PF00107; adh_zinc. 1.
DR Pfam; PF00109; ketoacyl-synt. 1.
DR Pfam; PF02801; ketoacyl-synt_C. 1.
DR Pfam; PF00550; pp-binding. 1.
DR Pfam; PF00975; Thioesterase. 1.
DR PROSITE; PS00075; ACP_DOMAIN. 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE. 1.
DR PROSITE; PS00012; PHOSPHOPANTHETINE. 1.
KW Phosphopantetheine; transferase.
SQ SEQUENCE 2439 AA; 257507 MW; D3A972BC49FC3497 CRC64;
Query Match 56.8%; Score 4098; DB 2; Length 2439;
Best Local Similarity 42.8%; Pred. No. 3.2e-192;
Matches 922; Conservative 139; Mismatches 315; Indels 776; Gaps 11;
QY 6 IERRADPIAIVAGASCRLPGVITLDSGFTLLSGSDTQGRVPAERMDAAWFDPPDPAP 65
DB 26 LQGERTEPIAIIIGIGRFGGADTPRAPFELDSGDAV--QPLDRMVLVGVHRESEVP 83
QY 66 GKTPTVTRAFPLSD-VACFDASFFGIGSPREALKMDPAHLLLEVCEWALENAIAPALVYG 124
DB 84 -----RWAGILTEAVGVGPAAPFGTSPREARSLDPOORLLLEVTEWGLBAGIAPOSLGG 138
QY 125 TETGVFIGIGPSEYEALPOATASAEIDAAGGLGTPSVGAGRIAYALGRGCVAVDPA 184
DB 139 SRTGVVLGACSSDYSTTVAAQORE-EQADYDITGNTLSVAAGRLSTLLGLQGFCLTVDA 197
QY 185 YSSSLVAHLACQSLRSGCESTALAGVSLMLSPSTLWLSKTRALARDGRCAFGSAD 244
DB 198 CSSSLVAHLACSLRSGCESTALAGVSLMLSKTMIMLGRIQALSPGHCTFDSAN 257
QY 245 GGRGGGCAVYVULKRLSGARADDRILAYTRGSAINHDSAGSLTPVNGSSQGITVLRAL 304
DB 258 GFVRGGCGMNVYVULKRLSDAQRHGRDIWALIRGSAMQDGRSTGLAMPVYLAQELRALRL 317

QY 305 ADAGCAASSGVVEAHGTGTTLGDPIEIOALNAVYGLGRDVATPPLIGSVKTNLGHPEYA 364
DB 318 QSAVADAGALDVEVHGCTGTSIGDPIEVDALRAVWGPARADSRCVLAGVKNLGLBEA 377
QY 365 SGITGLLKVLSLQHQIIPAHILHAQALNPRISWGLRLTVTRAPTPWPMNTPRAGVS 424
DB 378 ACVAGIKKAAALHHESIIPRNHFTLNPRIIEGTALALATEPVPMPAGRPFAVGA 437
QY 425 FGMSGTNHVVLEEA PAATCTPPAPERPELLVLSARTASALDAQAARLDLETYPSC 484
DB 438 FELSGTNHVVLEEA PATVLA PATGRSHELLVLSAKSTALDAQAARSAHIAIPEGC 497
QY 485 LGDVAFSLATTSAAEHRLAVALATRSREGIRALDAAQOTSPCAVRSIADSSRGTAFL 544
DB 498 LGDVAFSLVATRSPEHRLAVALATRSREARSLAEAAQOTPAGARGAASPGTARL 557
QY 545 FTGGAGTQLGMRGLYDVWSAFREAPDLCTVRLPNOELDPLREVMKAEPAVDALLDOT 604
DB 558 FAGGGAQVPGMGRGLMEAMPARETFDRCVTLFDRELHQPLCEVWMAEPGSSRSLLDQT 617
QY 605 AFTQPALTFEVALAALMRSGVEBELVAGHSIGELVAACVAGVSLIEDAVFLVARGRL 664
DB 618 AFTQPALTFEVALAALFRSGVEBELVAGHSIGELVAACVAGVSLIEDAVFLVARGRL 677
QY 665 MQALPAGAMVSI EAPEADVAAVAVPAHASVSI AAVNAPDQVVIAGAGQPVHAI AAMMA 724
DB 678 MQALPAGAMVSI EAPEADVAAVAVPAHASVSI AAVNAGEQVVIAGAEFVQIIAAAFPA 737
QY 725 RGARTKALVSHAFSPPLMA PML EAFGRVAESVSYRPSIVLVSNLSGACTDEVSSPGY 784
DB 738 RGARTKPLVSHAFSPPLMDPML EAFRRVTESVYRPSMALVSNLSGCRPCDEVCAPIY 797
QY 785 WVRHAREVRFADGVKALHAAGCTFEVGBSTLLGVPAQMPARPALSSRGRBE 844
DB 798 WVRHAREVRFADGVKALHAAGCTFEVGBSTLLGVPAQMPARPALSSRGRBE 857
QY 845 PATVLEALGGLWAGVLSMAGLPFGSGRRVPLPTYPMQREYWI DTKADDAARGRRA- 903
DB 858 AASALEALGFPVVGSGVTSWGSVFPSGGRVPLPTYPMQREYWI EAPVDSAGDQIGRAG 917
QY 904 -----PGAGH----- 908
DB 918 AGDHPILGEAFSVTHAGRLMETTLDRKRLPMLGEHRAQGVVFPAGYLMALSSGAE 977
QY 978 ILGDPIQVTVLLETLPAGDTAVPVQVVTTERPGRLPQVARSREBARAFRIHA 1037
DB 909 ----- 908
QY 1038 RGVLRVGRAEFPALNLAALRAHLAAVPAALITGALAEMGLQYGRALGLAELMRGEG 1097
DB 909 ----- 908
QY 1098 BALGRVLPESAGSATAYQLHVLVDACVQIVGAFADREBATPMAVPEVGSVRLFORSP 1157
DB 909 ----- 908
QY 1158 GELMCHARVSDGQAPSRWSADFELMDGTGAVVAEISRLVVERLSAGVRRRADDWFLE 1217
DB 1218 LDMEPALLEGKPTAGRMILLGEGGLGRSLCSALKAAGHVVAAGDDTSAAGRALILA 1277
QY 909 ----- 908
DB 909 ----- 908
QY 1278 NAFDQAPTAVVHLSLDGGQLDPLGAGALDAPRPVDADALLESALMRGCDVSL 1337
DB 909 ----- 908
QY 1338 VQALVGMDLRNAPRLMLLTGGAQAAAAGDVSVVQAPLLGLGRTIALBHAELRCISVDLP 1397

Oy	909	-----DEVEGGANVRGDRRSARLDHPPEPSGREKXEAGDPPRELEI	952
Dd	1398	AQPEGADALLAELIADDAAEEBVALRGERFVARLVHRLPBAQRREKTAPAGDRFFRELI	1457
Oy	953	DEPGVLDLIVLRVTERRAPGLGEYEIADAAGLSFNVDOLALGMVPDDLPGKNPPLLIG	1012
Dd	1458	DEPGVLDDQLVLRATGRARAPRGVEINVAANGLOSIDIQLANGVARNPLPGGELIEPVILG	1517
Oy	1013	GECAGRIYAVEGVNGVLVGOPVIALSAGAFAHTTSSAALVLPFPOLSAIEEAAMFVA	1072
Dd	1518	SECGARIYAIVEGVNGLVVGOPVIALAGVFPAHTVTTSATVLPPLDISATEAAMPDLA	1577
Oy	1073	YLTMWYALDRFARLQOPGRVLIHMAATGVGLAAVOMAQHNAEYNATACTGEKAYLESL	1132
Dd	1578	YLTMYALDKYAAHIQAAGERVLIIRAAAGSIGCAVMAGRVAAEVATADTEKRAYLESL	1637
Oy	1133	GVRVYSDSRSPFVADVAVMTGEGEVDVYLSLSELIDKSFNILRSHGREFELGKRDCY	1192
Dd	1638	GVRVYSDSRSGRFADVAHMTDGEVDVYLSLSEHIDKSLMTRACGRVLVKGRDDC	1697
Oy	1193	ADNOLGLEPFLRNLSFSLVLDLKGWMLERPARVRLLEEILGILIAAV-----	1239
Dd	1698	ADTOPGLPPLLRNFSSGOVDRKMWLDPARIRALLDELFGIVAAGALSIPSGILRVGS	1757
Oy	1240	FTTPPIALPLIARVADAFRSMQAQHCLKVLYTLGDDEVQVRIPIPHA-----	1286
Dd	1758	LTPPEVETFPISRAAEAFRRMAOGHLKVLTLTDDEPVIRIRAPAESSVAARADGTLYVT	1817
Oy	1287	-----GAG-----	1289
Dd	1818	GGLGGLGRVAGWLAERGAGQLTVLGRGAASABEQRAAVALAHGARVTAAKADVADR	1877
Oy	1290	-----	1289
Dd	1878	QIERVLRVITASGMPLRGVVAAGLVDDGLMQOTPARLTRVMGPKVQCALHLHTLTREA	1937
Oy	1290	-----	1289
Dd	1938	PLSFVLYASAAGLFGPSPOGNYYAANAFLDALSHHRAHGLPALSIDMGFEVGNAVA	1997
Oy	1290	-----PST-----	1292
Dd	1998	QENRGARILISGMRCITPDEGISALARLLLEGDRVOTGYIPTTPROWVEFYPTAASRLS	2057
Oy	1293	-----GDRDLLDRLASAAPAPAAAALAEALRTQVSQVLTPEIKYGAERALF	1338
Dd	2058	RLVTTQRAVADRRTAGDRDLLQGWSAESBSARAGLQDVVARQVSHVLALPBEDKILEVDAFL	2117
Oy	1339	TRLGMDSIMAVELENRRIRASLKLIKSTFLSTSPNIALLAONLL-DALATALV	1389
Dd	2118	SSMGWDSIMSELERNRIRIALLGVAPALAGTWYPTVAAITRWLLDDALAVALV	2169

RESULT 5

ID	09KI27	PRELIMINARY;	PRT; 7257 AA.
AC	09KI27;		
DT	01-OCT-2000	(TREMBLrel. 15. Created)	
DT	01-OCT-2000	(TREMBLrel. 15. Last sequence update)	
DT	01-MAR-2003	(TREMBLrel. 23. Last annotation update)	
DE	EPOD.		
GN	Polyangium cellulosum.		
OC	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;		
OC	Sorangineae; Polyangiaceae; Polyangium.		
OX	NCBI_taxid=56;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SMP44;		
RX	MEDLINE=20291058; PubMed=10831849;		
RA	Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.,		
RT	"Isolation and characterization of the epothione biosynthetic gene		
RT	cluster from Sorangium cellulosum.";		

[illegible]


```

Db 3501 ASLSLSTRSPMEHRLAIAITTSREALRGALDAAAOQOTPOGAVGKGVSSRGKLAFLFTQ 3560
Qy 549 GAOTLGMGSLYDVMSAPREAFDLCYRLFNQELDRLEVMMAEPASVDAAILDQAFPO 608
Db 3561 GAOVMGSGLEAMWAFREAFDRCALDRELDOFLREVMMAEPALQADQAFV 3620
Qy 609 PALFTFEVALAALWMSGVEPELVAGHSIGELVAACVAGFSLIEDAFLVAARGRLMAL 668
Db 3621 PALFTFEVALAALWMSGVEPELVAGHSIGELVAACVAGFSLIEDAFLVAARGRLMAL 3680
Qy 669 PGGAMVSTLEAPADVAAPVAAPHAASVSTIAAVNAPQOVIIAGAGQPVHAIAMAAARGAR 728
Db 3681 PGGAMVSTLEAPADVAAPVAAPHAASVSTIAAVNAPQOVIIAGAGQPVHAIAMAAARGAR 3740
Qy 729 TALAHSFHSPLMAPMLEAFGRVAVESYRPSIVLVSNSGKCTDEVASPGVMVH 788
Db 3741 TRLAVSHAFHSPLMAPMLEDFORVATTAIRAPDRPVSNVGHVAGPELIATPEYVH 3800
Qy 789 AREVVRPADGVYALHAAGAGTEVEVQKSTLLGLVPCMPDARPALASSRAGRDEPATV 848
Db 3801 VRSVVRFGDGAKALHAAGATFVEIGPKVLLGLLPACIGEADAVLVPBLRDRSCFV 3860
Qy 849 LEALGSLAVGSLVSWAGLPESGGRVPLPTYPQOERYWIDTKADA----- 896
Db 3861 LAALGTWYAMGALDMKGVFPDQARRVALPMYPMQERHMDLTPRSAPAGIAGRWP 3920
Qy 897 ----- 896
Db 3921 GVGLCMPGALVHNVSIGRHPDFLGDHLPFGKVVPGAFHVAVILSIAERWPERAIEL 3980
Qy 897 ----- 902
Db 3981 TGVEFLKATAMBDQVEVLAHVLTPPEAAGDGYLFELATIAETETERMTTHARGVQPTD 4040
Qy 903 -APGA----- 906
Db 4041 GAPGALPRLEVEDBAIQPLDFAGFLRLSAVRIGWGLMRLQDQGVGDEASLATVPT 4100
Qy 907 ---GHD-----EVEGGA-----VR-GGD 921
Db 4101 YPNADVAPLHPILLDNGFVALLATRSBEDDGTPLPFAVERVMWAPVGRVRCGV 4160
Qy 922 RRS-----RLDHP----- 932
Db 4161 PRSQAFVSSFLVDETGEVAVVEGFCVGRAPREVFLROESGASTAALYRLDWEBAIP 4220
Qy 933 -----ESGR 936
Db 4221 DAPARMESWVVAAPGSEMAAALATRLNRCVLAEPKGLEAALAGVSPAGVTCMBEPGA 4280
Qy 937 REKVAAGR----- 946
Db 4281 HEEAPAAQVATBGLSVQALRDRAVRLMWTTGAVAVEAGERVQVATAPVWGLGRV 4340
Qy 947 ----- 946
Db 4341 QERPELCTVLDLEPEVDARSADVLRLEGADETOVVFSGERRVAVKATTPBCL 4400
Qy 947 -----PFRLEIDPQVLDHVLVTRERRAPGLGEVEIAVDAAGLSFNDVOLALGNVPPD 1000
Db 4401 LVPDAESYRLBAGQKTDQLRLAPQRAPGPEVEIKVTSAGINFRVLAVLGM----- 4456
Qy 1001 LBGKNPRLGEGEGRIVAVGEGNGLVNGQPIYALSAGFAHHTVSALVLPKPPA 1060
Db 4457 YPGDAGP---MGDDAGVTAAGQGVHLSVGDVAVTTL---GTLHFPVYDABLVRQBP 4511
Qy 1061 LSAIEAAMPVAYVLTAWYALDRILARLOPGERVLIHAATGSGVLAIVQMAOHGAETHATA 1120
Db 4512 LTPQAATVPAFLTAMLALHDLGNRGERVLIHAAGSGVMAAVQIARWIGAEVPRATA 4571
Qy 1121 GTPERKAVYESLGV---RVVSDRSRDFVADVAPWATGEGVDVLSLSEGLIDKSFNLLR 1178

```

```

Db 4572 -SPSKAAVQAMGVBERTHIIASRTLEFAETPRQVOTGGRGVVVLNALAGEFVDASLSLT 4630
Qy 1179 SHGRFVEIGKRDQVADNQ---LGLRPELRNLSPLVDLGMMLERPARVRLIEELLGIA 1236
Db 4631 TGRFTLENGKTD-IDBRAVAAAHFGVRYFDLLEL-----ABDRTEILERVVEGFA 4683
Qy 1237 AGVTPPPIATLPIARVADAFRSMAOAOHLGKLVLTLDPEVOIRIPT-----HAGAP- 1290
Db 4684 AGHLALPVHAFIKAEAPAFPMQARHOGKVV-LAPSAAPLAPGTIVLTGGLAL 4742
Qy 1291 -----STGDRDLDRLASAAPAPAAALEAFLETOVSQVLRTEPIKVG 1333
Db 4743 GILHVARWLAQOQAPHVLTGRRG-LDTFGAKAVAEIHALGARVTI AASDVADRNALAEV 4801
Qy 1334 AEAL-----FTFLGMSLMAVBLRRIEASLTKXST 1365
Db 4802 LQAITAEWPLQGVIIAAGALDDGVLDDEOTTRFSSVLA PKVTGANNLHELTGNDLAFV 4861
Qy 1366 TFLSTS-----PNIALLAQNLDAATLALSERVAENLRACVOND 1406
Db 4862 LESSMSGLSGAGQSNVA-AANTFLDALAHRRAGLAAGSLAWGPWSD 4909

RESULT 6
ID 09L8C7 PRELIMINARY; PRT: 7257 AA.
AC 09L8C7;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Polypeptide synthase.
GN EPOC.
OS Polyangium cellulosum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Sorangineae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=So ce90;
RX MEDLINE=20130945; PubMed=10662695;
RA Molnar I., Schnupp T., Ono M., Zirkle R.E., Milanow M.,
RA Nowak-Thompson B., Engel N., Toupet C., Stritzmann A., Cyr D.D.,
RA Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;
RT "The biosynthetic gene cluster for the microtubule-stabilizing agents
RT epothilones A and B from Sorangium cellulosum So ce90."
RL Chem. Biol. 7:97-109(2000).
DR EMBL; AF210843; AAF26921.1; -.
DR HSSP; P25715; 1MLA.
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; FabD.
DR InterPro; IPR000794; ketoacyl-synt.
DR InterPro; IPR000612; peptidase_S8.
DR InterPro; IPR006162; Peptide_attach.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR002364; OOR_zeta_crystal.
DR Pfam; PF00698; Acyl_transferf; 4.
DR Pfam; PF00107; adh_zincf; 2.
DR Pfam; PF00109; ketoacyl-synt; 4.
DR Pfam; PF02801; ketoacyl-synt_C; 4.
DR Pfam; PF00550; pp-binding; 4.
DR TIGRFAMs; TIGR00128; fabD; 4.
DR PROSITE; PS00075; ACP_DOMAIN; 4.
DR PROSITE; PS00606; B_KETOACTL_SYNTHASE; 4.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
DR PROSITE; PS01162; OOR_ZETA_CRYSTAL; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 7257 AA; 764272 MW; 621A6820FA57ACE9 CRC64;

Query Match 45.9%; Score 3311.5; DB 2; Length 7257;
Best Local Similarity 41.1%; Pred. No. 5.2e-153;
Matches 784; Conservative 162; Mismatches 433; Indels 531; Gaps 25;

```

QY 9 AAEDPIAIVGASGRLEPGVITLDSGFMTLEGSRDYGVPAERMDAAWFDPDPAKGT 68
 DB 3021 ASBEPPIAIVGASGRLEPGVITLDSGFMTLEGSRDYGVPAERMDAAWFDPDPAKGT 3080
 QY 69 PUTRASFLSDVACFDASFFGISPREALRMDPAHRLLEVCMEALENNAIPASLVGTETG 128
 DB 3081 YVTKGAFRLDQGLDATTFFRISPREAMSLDPQGRLLLEVMELASGIAIPDLRDSPTG 3140
 QY 129 VFIIGIPSEYEALPQATASAEIDAGGLTSPSVGAGRISYALGRPCVAVDTAYSS 188
 DB 3141 VFVGAAPNEYTYOTRLRFTDGAAGLYGGTNMLSVYAGRLSFTGLGPTLMDTACSS 3200
 QY 189 LVVHLACOSLRSGECSGTALAGVSLMLSPSTLWISKTRALARDCKAFSAADQFGR 248
 DB 3201 LVVHLACOSLRSGECSGTALAGVSLMLSPSTLWISKTRALARDCKAFSAADQFGR 3260
 QY 249 GEGCAVVLKRLSGARADGRIIAVIRGSAINHGAASSGLTPNGSSQELVKRALADAG 308
 DB 3261 GEGCAVVLKRLSGARADGRIIAVIRGSAINHGAASSGLTPNGSSQELVKRALADAG 3320
 QY 309 CAASSGVYVAGTGTTLGPIELQALNAVYGLGRVATPULLGVSYTNIGHBEVYSGIT 368
 DB 3321 VSEVDVDFVECHGTGTALGDIYEQALSEYVGRSGDRPLVGAAGKAVAHLEAASGLA 3380
 QY 369 GLTKVVLISQHGQIPAHILHAQALNPRIISMGDLRLTVTRARTPMPDMNTPRAGVSPGMS 428
 DB 3381 SLIKAVIALRHEQIPROPELGEINLPHLPMNTLPVAVRKAIVPMGRKARPRAGVSAFGLS 3440
 QY 429 GTNAHVLEEAAPACTTPRAPERPAELVLISARTASALDOAARLRDLHETVPSQCLGV 488
 DB 3441 GTNAHVLEEAAPACTTPRAPERPAELVLISARTASALDOAARLRDLHETVPSQCLGV 3500
 QY 489 ABEFLATTRSMERILAAVATSREGRLAALAAAGQGSPPAVSISADSSGKLAFLETCQ 548
 DB 3501 ABEFLATTRSMERILAAVATSREGRLAALAAAGQGSPPAVSISADSSGKLAFLETCQ 3560
 QY 549 GAOTLGGGGLYUWMAAFREAFDLCVLFNOELDRPREVMMAPASVDALLDQTAFTQ 608
 DB 3561 GAOTLGGGGLYUWMAAFREAFDLCVLFNOELDRPREVMMAPASVDALLDQTAFTQ 3620
 QY 609 PALFTEYALAAALWRSWGVPELVAGHSIGELVAAVAGVFSLEDAVFLVARGRLMOAL 668
 DB 3621 PALFTEYALAAALWRSWGVPELVAGHSIGELVAAVAGVFSLEDAVFLVARGRLMOAL 3680
 QY 669 PAGAMVSIAPREPDVAAAAPHAASVSIANAAPDOVITAGACQPHALIAAAMAAARGAR 728
 DB 3681 PAGAMVSIAPREPDVAAAAPHAASVSIANAAPDOVITAGACQPHALIAAAMAAARGAR 3740
 QY 729 TKALHVSHPPLMAFMLEAFGRVSAESVYRRPSIVLVSNLSGKACTDEVSFGYVVRH 788
 DB 3741 TKALHVSHPPLMAFMLEAFGRVSAESVYRRPSIVLVSNLSGKACTDEVSFGYVVRH 3800
 QY 789 AREVVREADVAKALHAAGAGTVEVGPKSTLLGLVPCMPDARPALLASSRAGDEPATV 848
 DB 3801 AREVVREADVAKALHAAGAGTVEVGPKSTLLGLVPCMPDARPALLASSRAGDEPATV 3860
 QY 849 LELGLGIMAVGLVSNAGLEFPGSGRRVPLPTYMQRRRYWITDKAADA 896
 DB 3861 LELGLGIMAVGLVSNAGLEFPGSGRRVPLPTYMQRRRYWITDKAADA 3920
 QY 897 896
 DB 3921 GVGICMGAVLHVHVLISGRPHQPLGPHLVPGKVVGCAHVAIVLISIAERWPERAIEL 3980
 QY 897 902
 DB 3981 TGVEFLKAIAMEPQVEVLAHLVLPBAAGDYLFEELATLAETERMTTHARGVQPTD 4040
 QY 903 -ARGA- 906
 DB 4041 GARGALPRLLEVEDRAIOPIDFAGFLDRLSAVRIGMWPLWLODGRVGDASLTLVPT 4100

QY 907 ---GHD-----EVEBGA-----VR-GSD 921
 DB 4101 YENAHVVAHLRHLILDNGFAVSLSTRSEPEDGTPPLPFAVERVMWRAPVGRVRCGV 4160
 QY 922 RNSA-----RDHP 932
 DB 4161 PRSQAAGVSSPVLVDETEGVVAEVEGFCRARAREVFLRQESGASTAALYRLDWEAPAPL 4220
 QY 933 -----ESGR 936
 DB 4221 DAPAEIESKVVVAAPGSEMAAALATRLNRCVLAEPKLEALLAGVSPAGVICTMEAGA 4280
 QY 937 REKEVAADR----- 946
 DB 4281 HEEAPAAQRAVATBGLSVVQALRDBAVRLMWVTVGAIVAEAGBRVQVATAPVWGLGRVYM 4340
 QY 947 ----- 946
 DB 4341 QERPBLCTVLDEPEADARSADVLLRELGRADDETOVAFRSGRKRVARLVKATPEGL 4400
 QY 947 -----PRLTIDEPGLDHLVLRVTERAPGLGEVEIIVDAAGLSFNDVOLALGMVPPD 1000
 DB 4401 LVPDASTYLEAGQKTDQULAPAKORAPGEGVEIVTASGLNFRIVLAVLGM----- 4456
 QY 1001 LPCKPNPULLGECAGRIIVAGVEGVGLVGOPIYALSAGAFATHVTTSAALVLPQGA 1060
 DB 4457 YEGDAGP---MGDDGAGVATAVGGQVRHVAVDAAVMTL---GTLHFRVTVDAELVVRQAPAG 4511
 QY 1061 LSAIEAAMPVAVLTWVYALDLRIARLOPBERVLINAAATGCVGLAANVQMAOHGAEVHAHA 1120
 DB 4512 LTPQAATVPVAFLTMTMLHDGLNRGERVLIHAAAGGVMAAVQIARWIGAEVFATA 4571
 QY 1121 GPBEKAVLESIGV---RYVSDRSDFVADVAMTGEADVVLNSLSGELIDKSFNLR 1178
 DB 4572 -SPSKMAVQANGVPRTHIASRTLEFAETPQVOTGREGVDVLAALAEFPDASLSLS 4630
 QY 1179 SHGRFVELGKRCYADNO---LGLRPELRNLSFSLVDLRGMLEPARVALLLEELGLIA 1236
 DB 4631 TGRFLEMGKTD-IRBRAVAAAHPCVRYRFDIEL-----APDRTEILERVVEGPA 4683
 QY 1237 AGVFTPPPLATPIARVADAFRSMAOHLGKLVLTGDPVOIRPT-----HAGAGP- 1290
 DB 4684 AGHTRALPVHAFATTAENAAFRFMAQARHQGVVL-LPPPSAAPLAFTGTVLLTGGLAL 4742
 QY 1291 -----STGDRDLDRLASAAPARAALAEFLRTQVSGVLRTPETIKVG 1333
 DB 4743 GLHVAWMLAQGVPHAVLTGRG-IDTPGAALAVAEIETALGARVTTAASDVADRNLAEV 4801
 QY 1334 AEARL-----FTRLGMDSLMAVELRNRIEASLKJLKJST 1365
 DB 4802 LQIIPAEWPLQGVIIHAAGALDGVLEQTTDRFSRVLAPEKTVGAMWLHETAGNDLAFV 4861
 QY 1366 TPLSTS-----PVALAONLIDALATALSLEVAENLRAGVQND 1406
 DB 4862 LFSMSGLGSAGQSNYA--AAVTFLDALAAHRRAREGLAAGSLAWGWSMD 4909
 QY 1406

RESULT 7
 Q8RUY1 ID Q8RUY1 PRELIMINARY; PRT; 2218 AA.
 AC Q8RUY1;
 DT 01-JUN-2002 (TREMUREL. 21, Created)
 DT 01-JUN-2002 (TREMUREL. 21, Last sequence update)
 DT 01-OCT-2002 (TREMUREL. 22, Last annotation update)
 DE Scif protein.
 GN STIF.
 OS Stigmatella aurantiaca.
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
 OC Cystobacterineae; Cystobacteraceae; Stigmatella.
 NCBI_TaxID=41;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SG a15;

RA Gaitatzis N., Siłkowski B., Kunze B., Nordstiek G., Blocker H.,
 RA Hofle G., Muller R.;
 RT "The biosynthesis of the aromatic myxobacterial electron transport
 RT inhibitor actinobactin is directed by a novel type of modular
 RT polyketide synthase";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A421825; CAD19090.1; -
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR002085; Adh. zn family.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR006153; Pp_bind.
 DR Pfam; PF00698; Acyl-transfer; 1.
 DR Pfam; PF00107; adh_zinc; 1.
 DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF02801; ketoacyl-synt; C; 1.
 DR Pfam; PF00550; pp-binding; 1; 1.
 DR PROSITE; PS50075; ACP DOMAIN; 1.
 DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 1.
 DR Phosphotransferase; Transferase.
 KW SEQUENCE 2218 AA; 237399 MW; B85339D5457FA4F5 CRC64;

Query Match 36.5%; Score 2634; DB 2; Length 2218;
 Best Local Similarity 35.4%; Pred. No. 1,7e-120;
 Matches 642; Conservative 182; Mismatches 447; Indels 542; Gaps 25;

8 RAEDPIAIVGASCRPLGVDLISGFWTLLEGSDTVGRVPAERMDAAAFDPDPAPGK 67
 24 RLAPETIAIVGMAKRPFGVKSIDSYNDLLKRGDVAITEIPARMDINAFYDPNDSPEK 83
 68 TEVTRASFLSDVACFDASFFGISPREALRMDPAHRLLEVCWALEENAIAPASLVGTET 127
 84 TTRMCGFLDVEGFDPHLFGISPREASMDPHQRLFLVAMEALEENACQDTRLAGSKT 143
 128 GVFITGIGSEYE-AALPOATASAEITAHGGLGTMSEVGAGRISYALGLGPCVAVDTYS 186
 144 GVFVGLYTSIDYSPVVLSDNA--INVYGTGTANNVAAAGRLSYLLDQPSFVDTACS 200
 187 SELLVAHLACOSLSRSGESTALAGVSLMSPETLVMLKTRALARDGRCKAFSAEDGF 246
 201 SELLVAHLACOSLSRSGESTALAGVSLMSPETLVMLKTRALARDGRCKAFSAEDGF 260
 247 GREGCAVVVLRKLSGARDGRIILAVIRGSAINHDSAGSLTPVNGSSQEIIVLRKALAD 306
 261 ANEGCGGVVILKRLSDPADGRIILAVIRKASAINQDGRITNGLTAFSGASQVILRQALTR 320
 307 ACCAASSVGVFAHAGTGTLGDPTEIQALNAVYGLGRDVATPLIISVYTNIGHEPYAG 366
 321 AGISPSQVTVFAHAGTGTLGDPTEIQALNAVYGLGRDVATPLIISVYTNIGHEPYAG 380
 367 ITGLKLVVLSLOHGOI PAHLHAQALNPRI SWGDLRLTVTRARTPMPDMNTPPRAGVSFG 426
 381 VAGLIVVLSMOKEALIPOLHRELNSNISLENTSLVTRKELHPWAGCGPRIAGISAG 440
 427 MSGTNAVHLEAPATCTPPAPERPAELVLISARTASALDAQAPALRDLHLETYSOCLG 486
 441 MSGTNAVHLEAPATCTPPAPERPAELVLISARTASALDAQAPALRDLHLETYSOCLG 500
 487 DVAFSLATTRSAMERFLAVATSRBGLRALDLAAOGQTSFGAVRSIADSSRCK-LATLF 545
 501 DLYSANTGMSKFEHLIAPAGTIAQLSRLSFLTEGQCPQAGTSYQVTVGRGPGVAFIF 560
 546 TGGGAGTLLGKGLVYVMSAFREAPFLCVRLFNQELDRPLREVMMAEPASVDAALLDQTA 605
 561 TGGGAGTLLGKGLVYVMSAFREAPFLCVRLFNQELDRPLREVMMAEPASVDAALLDQTA 618
 606 FTQPALFTTEYALAAALMRMGVEPELVASHISIGELVAVACVAGFSLIEDAVFLVAARGRL 665
 619 FTQPALFAEMALAEIMRSMGVPTVVMGSHVGEYVAAACVAGFSLIEGALILVERARLM 678
 666 QALPAGGANVSI EAPADVAALVAAPHAASVSTAAVAPDOVVTAGGQVYHAIAAAMAR 725
 679 QSLPBGEMAAYFTDQARTAOAIAPASQVSI AAFNGPSETVYSGDGKAVEAILGALSAE 738

QY 726 GARTALVSHAFHSPLMAPMLEAFGRVAESVYRPSIIVLSNLSGACTDEVSPGWM 785
 DB 729 GYKVRRLGSHAFHSPLMDPMLDAFERAAKTRFMPKTTILSNLTG-GPSEFSAAL- 796
 QY 786 VHAHEVVRFPADGVKALHAAGACTVEVGPXSTLLGLVPACMPDARPALASSRGRDEP 845
 DB 797 RHAREPVRFSFGMAALARGVVAIVEIGPNATLLGISRCLPESVAMVPLSRDKDEM 856
 QY 846 ATVLALGLMAVGLVSMAGL--FPGSGRVPLPTYPWQBERY----- 887
 DB 857 EVLLATTLGLSLVGRVPVWAGDSYPR--RRVSLPNTPFQHERALKOMDGRGVAIPSCQ 914
 QY 888 ----- 887
 DB 915 VVGHPLLSRLRVATVDGTLFESKFRADBPSPFSEHRVYGMSTVPATVLEGALAAEBI 974
 QY 888 ----- 887
 DB 975 FGAGHALESVDIOEALVLSDAPTVOYLIGSREBGTARFQVPSIASSGSGEGWTTTH 1034
 QY 888 ----- 890
 DB 1035 SSGQIRIASGAPPLSGTLPTEQALEAIRTRCPGLEPRAFYDWLEGGOLEYGRFRGQV 1094
 QY 891 ----- 912
 DB 1095 QVMRGGDALGLVELPEBAVDKAKFKVHPGLMDACVQLFGVIEYREGAODTEVYLPVS 1154
 QY 913 ----- 914
 DB 1155 IDRYLKGRLGSKVSHARIRPSEGTARETLKADIRLYDEAGRIVAEEGMCIKRAPREM 1214
 QY 915 ----- 928
 DB 1215 IARFQAOFSMLHEWKEEYRASAAGAARGSCRWVAFDRSGVABALIDRLTAGRP 1274
 QY 929 ----- 936
 DB 1275 VVRVLADHYSREGVLRLHPQEPHVRSLMESASGSAVTRVYLWGLDSSGVSNGE 1334
 QY 937 REKVE--AAGDRP----- 947
 DB 1335 ASRVDTRAIGSNPLHLVQELAGQGSQHLPLVTRGAQATTREHAPIDAFQASLWGFAAT 1394
 QY 948 ----- 947
 DB 1395 IAFHEPELRPVRIIDLEPSRKSGSPRETEILGSEVTRSDASAHDAIAFRGSLRLVSLMR 1454
 QY 948 ----- 989
 DB 1455 QAKRASNRKRLRIPAPAPFRLEISQRTLEDTLVAPVERRRPQPEVEIRVAPATGLNFRD 1514
 QY 990 VQALGMPVDDLPGKPNPPLLIGCECAGRIVAVEGVNGLVVGQVIA-LSAGAFATHT 1048
 DB 1515 VLNALGMPVGD-----PGLIGECACVISAIGVGEVGFKVGDPVVAVISGSAFYCT 1567
 QY 1049 TSAALVLRPQALSAIEAANPVAIVLTAMVALDRARLQPERVULIHAATGVGILAAQVM 1108
 DB 1568 VLADVVAHKRPALSPAOAAAIPIAELTAOYGLQOIGRMAAGSRVLIHAAGGVGAAVQL 1627
 QY 1109 AQHGAEVATATAGTEPKAVYLESLGRVYVDSRSDRFADVAAMTGGEGVDVNLSSGE 1168
 DB 1628 AKRGADEVFATDSS-GKMDVLEAMKVERMRMSRTIDFADVMAKREGGIDIVLMSLAD 1686
 QY 1169 LIDKSFNLIRSHGFELGKDCVADNQLGRPLRLNLSFSLVLDLKGMLERPARVALL 1228
 DB 1687 FIPKSLVYLSRSGRELIGKGVWTSAGVAER--YPGVSVAIYDLGVTXOPSIRSMF 1744
 QY 1229 BEILGLIAGVTPPPIATLPIARVADAFRMAQOHGKVLTLGD--PEVQIITPTHA 1286
 DB 1745 PALMFEFGTILVPPPLLVFALQDAVEAFRHMAALAKHGVKVVITVNDGAP-SAETARA 1803
 QY 1287 GAG-PTGDRDLL 1298

Db 1804 GAAVPIRGDSTYL 1816

RESULT 8

Q9ADL6 PRELIMINARY; PRT; 6315 AA.

AC Q9ADL6; 01-JUN-2001 (TREMURE1.17, Created)

DT 01-JUN-2001 (TREMURE1.17, Last sequence update)

DT 01-MAR-2003 (TREMURE1.23, Last annotation update)

DE Sorafen polyketide synthase A.

GN SORA.

OS Polyangium cellulorum.

OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;

OC Soranginaceae; Polyangiaceae; Polyangium.

OX NCBI_taxid=56;

RN 1)

RC SEQUENCE FROM N.A.

RC STRAIN=So ce26;

RA Ligon J.M., Hill S., Beck J., Zirkle R., Molnar I., Zawodny J.,

RA Money S., Schupp T.,

RT "Characterization of the biosynthetic gene cluster for the antifungal

RT polyketide sorafen A from Sorangium cellulosum So ce26."

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES

CC (SDR) FAMILY.

DR EMBL: U24241; AK19883.1; -

DR HSSP; P25715; 1MLA.

DR InterPro: IPR001227; Ac transferase.

DR InterPro: IPR002198; ADH_short.

DR InterPro: IPR002085; Adh zn family.

DR InterPro: IPR004410; Fadh.

DR InterPro: IPR000794; Ketocacyl-synt.

DR InterPro: IPR006162; ppancne_attach.

DR InterPro: IPR006163; Pp_bind.

DR InterPro: IPR005823; Sug transporter.

DR Pfam: PF00698; Acyl transfer; 4.

DR Pfam: PF00106; Adh_short; 3.

DR Pfam: PF00107; adh_zinc; 2.

DR Pfam: PF00109; ketocacyl-synt; 3.

DR Pfam: PF02801; ketocacyl-synt_C; 3.

DR Pfam: PF00550; Pp-binding; 4.

DR TIGRFAMs: TIGR00128; fadh; 4.

DR PROSITE: PSS0075; ACP DOMAIN; 4.

DR PROSITE: PSS0012; B KETOACYL SYNTHASE; 3.

DR PROSITE: PSS0012; PHOSPHOPANTETHEINE; 3.

DR PROSITE: PS00216; SUGAR TRANSPORT 1; 4.

DR Oxidoreductase; Phosphopantetheine; transferase.

KW

SEQUENCE 6315 AA; 671301 MW; 783FB977006C12FE CRC64;

Query Match 36.1%; Score 2606; DB 2; Length 6315;

Best Local Similarity 35.8%; Pred. No. 1.7e-118;

Matches 665; Conservative 193; Mismatches 489; Indels 508; Gaps 32;

QY 10 AEDPIAIVGASCRPLPGVVIDLSCGFVTLLEGSRDVTGVPARWMDAAMPDPDPAPGKTP 69

Db 2091 SDEIRIALVGMALRLPGSIGVDALMDLHGRDVERIPIPRMDAGLYPDPDPARKKSY 2150

QY 70 VTRASFLSDVACPDASFFGISPREALRMDPAHRLILEVCWPALENAAIADPSALVGTETGV 129

Db 2151 VRHAMADQVDLFDPGFFGISPREAKHIDPQHRLLLEAAMQALEAGIVSTLKDPSPTGV 2210

QY 130 FIGGSEYEALPQALPASAEIDAHGGLGTMPSVGARISYALGLRCPCAVDTAVYSSSL 169

Db 2211 FVGIGASEY--AREPGAEDESEAYIVQGYASFAARLLAFTLGLQGPALSVDPACSSSL 2267

QY 190 VAVHACQSLRSGCESTALAGVSLMSPSTLVLSKTRALARDGRCKAFSAEADGFGRG 249

Db 2268 VALHLAGALRDRDCNLALAGASVMSPTFTVLSLRLALAPGRSKTSASADGIGRG 2327

QY 250 EGCADVVLKRLSGRADGDRILAVIRGSAINHDGASSGLTVPNGSGQETIVKRALADGC 309

Db 2328 EGVTVALERLNDALAQGRVLAVVGTAVNHDGSSGITAPNGTSQKKVLAALHDARI 2387

QY 310 AASSVGVFAHGTGTLDPIELIQALNAVYGGRDVATPLIGSVKTNLGHPEYASGITG 369

Db 2388 APADVDVVECHGTGSLGDPLEVOALNAVYGGERSAEKELFLGAVTIVNGHLEAAGLNG 2447

QY 370 LKRVLSLHGOI PAHLHAQALNPRIISWGLRLVTYRATPWPDW--NTPRDAVSSFGM 427

Db 2448 VAKIVASLHNLNLPPLHTTPRNPPLIAMDALAVAVDVAITRPVNRHADGRPRRAGVSAFGL 2507

QY 428 SGTNAVVEEAPAACTPPAPERPAE-----LLVLSARTASALDAQALRDHLETY 480

Db 2508 SGTNAVVEEAPAACTPPAPERPAE-----LLVLSARTASALDAQALRDHLETY 2567

QY 481 PSQCLGDVAFSLATTSAMEHRLAVATSRGRLAALDAAGQOTSPGAVRSIADSSRGK 540

Db 2568 SELALADVAYSLATTSAMEHRLAVATSRGRLAALDAAGQOTSPGAVRSIADSSRGK 2624

QY 541 LAFLFTGGAQTLMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMABEASVDAL 600

Db 2625 LAVLFTGGSQRTMGRALYDAFPVFRDALDTVAHLDRDLDRPLRDVLFAPDGSQOAR 2684

QY 601 LDQTAFTQPALFTFEYALALNRSGVPELYAGHSIGELVAAACVAGVSLDEAVLVAA 660

Db 2685 LDQTAFTQPALFTFEYALALNRSGVPELYAGHSIGELVAAACVAGVSLDEAVLVAA 2744

QY 661 RGRIMQALPAGGAMVSEIAPADVAAPVAPHAASVIAAVNAPDQVVIAGOPVHAIA 720

Db 2745 RAKLMQALPAGGAMVSEIAPADVAAPVAPHAASVIAAVNAPDQVVIAGOPVHAIA 2804

QY 721 AMAARGARTKALVHSHAFHSPLMAPLEAFGRVAESVSYRDSIVLSNLSGKACTD-EV 779

Db 2805 QAEALGRKTRRLVSHAFHSPLMAPLEAFGRVAESVSYRDSIVLSNLSGKACTD-EV 2864

QY 780 SSGFYVVRARAEVVRADGVKALHAAGAGTFVEVGKSTLLGLVPRCMD----ARPALL 835

Db 2865 ASPDYVVRARAEVVRADGVKALHAAGAGTFVEVGKSTLLGLVPRCMD----ARPALL 2924

QY 836 ASSRAGRDEPATVLEALGGLMAVGLVSNAGLF-PESGGRVPLPTYPMGREERYMID-TRA 893

Db 2925 PTLRKRDDEAETALGALHAAGLTPDMSAFFAPAPAKVSLPTAFQREKRWLDASGA 2984

QY 894 ----- 893

Db 2985 HAADVASAGLSTHDLGAGVPLADRDGFLFTGRSLSEHPWLADHVFGTPILPGTAF 3044

QY 894 -----DDA----- 897

Db 3045 LELALFVAGRVGLDVEBELTETPLALPSEGALLVQVSVGPLDDAGRRPLSLHSPQOAP 3104

QY 898 ----- 897

Db 3105 QDAPTRHASGLAPATPSPBFDLHWPSPSGATQVDTQGLYATLESAGLAPGQFOGLAS 3164

QY 898 ---RGDR----- 901

Db 3165 VMRSGDELFAEAOQPDAAKKAARFALHPALLDSALHALALDDEPARGVALLPFSMGVSL 3224

QY 902 -----RARGGHDE-- 910

Db 3225 RAVGATTLRVFRPRPKGTAGSLVLDAAGPIASVOALATRTISAEOURTPASHHDAL 3284

QY 911 -----VEEGAVRG-- 919

Db 3285 FRVDMSELSPSPSGAPSAVNLGIGDLAPVPLARVADLAALDQASLPFGGLV 3344

QY 920 ----- 919

Db 3345 VPEMATADDLIGASISITABALALQAWLADERLASSRLVLLTRBAIARADEVKDIA 3404

QY 920 ---GDRSARLDHP-----PESGGRKEVAAAGDPF----- 948

Db 3405 HAPLWGLASASQSEHPELPLFLVDLDSLSEASQHTLLALETGRHS-RLRNGKPFILRLA 3463

QY 949 -----RLEIDEPVLDHLVL-RVTERRAP-GLGEVEIAVDAAGLSFNDV 990
 DB 3464 NARSDIELIAPDASNMRLHIPTKGNFDALTLVDAPLARAFLHGVAVVAHAANFNDV 3523
 QY 991 QALAGMVDLDPKKNPRLGCEGACGRIVAVEGVNGLVGCOPYIALSAGAFATHTTS 1050
 DB 3524 LDTLGL-----YFGDAGP---LOGEGAGIVTEVGPVSRTYVDRVMGIFGAACGPTALAD 3576
 QY 1051 AALVLRPOALSAIEAAMPVAYLTAMVYLDRIARLQGERVLIHAAGVGGLAVOMQ 1110
 DB 3577 ARMTCPPIPHANSFAQASVPITLYLTYGLVDLGHKKNQVLIHAAGVGGLAVVQIAR 3636
 QY 1111 HVGAENVHATAGTPEKRAVYLESIGV--RYVSDRSRDFVADVAWMTGEGEVVDVLSLGE 1168
 DB 3637 HLGAEVFNPA-SAGKMSALRALGFPDDAHLASRDLDFEHFLRSTHGVGVVDVLDCLARE 3695
 QY 1169 LIDKSFNLIRSHGRVVEIGKRDQVADNQLGLR-PFLRLNLSFSLVLDRCMMLERPARVAL 1227
 DB 3696 FVDAFLRLMPSCGREGVEMGKTDIRREPDVAVGAVPGVVRAFDLIE-----AGPDRIEQM 3749
 QY 1228 LEEELGLIAGVFTEPPRIATLPIARVADAFRSMAGOHGKULVTLG---DEVOQIRIPT 1284
 DB 3750 LAELSLBERGALRPPITSMDIRHAPQAFRALADRHGKFLVTIPRIDEGTVLI-- 3807
 QY 1285 HAGAPSTGDRDLDRLASAAPARAALAEFLRTOVSQLRTPEIKVGAELFTRLGMD 1344
 DB 3808 -----TGCTGTGALVARNHVARHGAHLLTSTR--QGANAP---GAEN--SRTELE 3852
 QY 1345 SLMA-VEIR--NRISASLKLKSTFTLSTSPNIAL--AQNILDALATLSIERV 1394
 DB 3853 ALGASVTLRACDAADPRALQALLDSIPSAHPLTAVVAHAAGLDDGLGMSPERI 3907

RESULT 9

Q8GBX6 PRELIMINARY; PRT; 5017 AA.
 AC Q8GBX6;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Polypeptide synthase.
 GN SPH.
 OS Polyangium cellulosum.
 OC Bacteria; Proteobacteria; Delicatoteobacteria; Myxococcales;
 OC Sorangineae; Polyangiaceae; Polyangium.
 CX NCB1_Taxid=56;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Knauer J., Bioecker H., Mueller R., Nordstiek G., Beyer S.;
 RT "The spirangliene synthase from Sorangium cellulosum 50 c690 - Module
 RT duplications as strategy behind the evolution of polypeptide
 RL synthases";
 RL Submitted (Aug-2002) to the EMBL/Genbank/DBJ databases.
 RL EMBL; AJ505006; CAD3449.1;
 SR SEQUENCE 5017 AA; 52555 MW; 08BED56182669493 CRC64;

Query Match 33.5%; Score 2416.5; DB 2; Length 5017;
 Best Local Similarity 39.8%; Pred. No. 2,4e-109;
 Matches 639; Conservative 186; Mismatches 521; Indels 259; Gaps 41;

QY 9 AADPIAIVGASCRPLGCVIIDLGSFMTLLGSRDTVGVAPARMWDAAMFDDPDAPGKT 68
 DB 1586 ASDEPIAIVGVGLRPLGVDVDDALFRLLGGRDAVAIPASRMVGVAVFDDPDPAKXT 1645
 QY 69 PYTRASFLSDVACPDASFFGISPREALMDPAHRLLEVCWALENAAIAPSAVGTETG 128
 DB 1646 YVRHAMLDKRYDLDFAGFEFGISPREARVYDPQHRLLLETAMQALBAATVPSASLDSTTG 1705
 QY 129 VFIGGSEVEEALPQATASAEIDAAGGLGTMPSGAGARISVALGRPCVAVDTAYSSS 188
 DB 1706 VFVGAGASDY--AVLOSSAE--DAEAYAAAGTAASPAAGRLATLGLQGPALSIDTACSS 1762

QY 189 LVAVHLAGQSLRSGECSTALAGVSLMSPSTLYMLSKTRALADGRCKAFSAEDGFCR 248
 DB 1763 LVALHLACQSLRSGECSTALAGVQVMASEPVFVILSTRALADGRCKTFSANADGFCR 1822
 QY 249 GEGCAVVLKRLSGARADGRILAVIRGSAIINHGCSSGLTPVNGSSGEIYLKALADAG 308
 DB 1823 GEGVAVLLELRBARANGHFLVAVRGTAVNHGASSGIRAPNGSSQOKYLRALADAR 1882
 QY 309 CAASGVVEAHGTTGLGDIETIOALNAVYGLRDVATPLIGSVKTNLGHPEVASGIT 368
 DB 1883 LAPVDVVECHGIGTSLGDIETIOALNAVYGEGRAPRPLRGAVKTNIGHLEVASGLA 1942
 QY 369 GLKLVVLSLQGOIPAHHLAQAALNPRIISMGDLRLVTTRARTPMDW--NTPRRAGVSSFG 426
 DB 1943 GVVKIVASLRKGTLPATHTMPPRPHIDMDALPRVVDAAAPWESSADGSRBRAGVSAFG 2002
 QY 427 MSGTNAAVVEEAPAAITCTPPAPERPELL-----VLSARTSALDAQALRLDHL 478
 DB 2003 LSGTNAHYIVEAP---EPDAPKARGALPLSGALPVLVLSAKSDAALRAQAALRLDGLA 2058
 QY 479 TYPGCGDVAFSLATTSSAMEHRLAVAATSRREGIARAALDAAGQTSRGAVRSIADSSR 538
 DB 2059 RTPDALVDVAALATITRSQFDHRAIADSHGLVALBGLAGVAAPGT--VAKGAA 2116
 QY 539 GKLAFLFTGQAGQTLGMRGLDYVWSAFREAFLCVRLFNQELDRPLREVMMAEPASVDA 598
 DB 2117 DKLAFLFTGQAGQAAAGRGLYDAFPVPRDALDAVASHLDBELDRPLRDVLFAPKGSILA 2176
 QY 599 ALLDQTAFTOTALTFEYALALMRSGVEBELVAGHSIGLVAACVAGVSLBDAYVLV 658
 DB 2177 SLDRTEFTQALPALLEVALFRLVEAMGVTDVILGHSVGLAAAHVAGVISTLDACTLV 2236
 QY 659 AARGRLMOALPA--GGAWVSIAREADVAAYAAAPHAASVIAAVAPDOVAVAGQPVHA 717
 DB 2237 AARARLMOALPARQGANVTYATENEVLALBEGDGAELIATINAPSSVTAGVDVALR 2296
 QY 718 IAAAMAARGARTKALVHSHAFHSPLMAFMLAFAGRVASEVSYRPSIVLVNLSGKACTD 777
 DB 2297 VAAHFEARGKATRLVHSHAFHSHPMDLDAFPRVAGLFFHPRIPITVSNVGRLLAD 2356
 QY 778 EVSSFGIVVHAREVVRPADGVKALHAGACTFVEVGPKSTLGLVPAKMPDAR--PALL 835
 DB 2357 EIRSPDYVVRHRSVAFADINTLEADGVSSFELGPHGVLSLGSALNAQREVAFV 2416
 QY 836 ASSAGDEPRTVLEALGGLVAVGVISWAGLF--PSGGRRPVLPPTYMQRERYMIDTAD 894
 DB 2417 PALRDGRADVDAITLAUSSLHVHGRVDMAAFEGFPDQVATLYAQRDRFMDADA- 2475
 QY 895 DAARGDRAPGAGHD-----EVEEG--GAVGCGRRSARLDH-----PPESGREGX 939
 DB 2476 -GATRDEAPAPAPBDAAFWRAVDAGVGL--GATLNASGEBHLGALATLPLALSAMRRAR 2533
 QY 940 VEAAGDRPFLIEDPGLVHLVLRV-----TERRAPGLGEVEIAVDAAGLS---ENDVQ 991
 DB 2534 -----DESLVDALRYVWKKPLTPPAADVAGTMTLVTLAAGADALARELE 2580
 QY 992 LALGMVDDLPKKNP-----LLGCEGACRIYAVEGVGVGLVVGQVITLSAGAFTH 1046
 DB 2581 RALTAGAEVETWPIAPBEADRVLAVRVQALADAG--LKGIV---SPALDETTLATHT 2636
 QY 1047 VTTSA--ALVLRPOALS--AIEAAMPV-----AYLTAV----- 1077
 DB 2637 TALRAGLALTLRLVQALGDKGIEAPLMLVTRGAVSTGRSDSLAAGSMTGLGRVYGLE 2696
 QY 1078 -----VALDRIARL--QPE-----RVLIHA----- 1096
 DB 2697 HPERWGLVDVADGDTIDARLDRVLTLASRDEBDOLALRPTGI PARRLVRAFLGAPR 2756
 QY 1097 -----ATGVC-----LAAYQAOH--VGAENVHATAGPEKRAVLESIGV 1134
 DB 2757 ARTYTAGTSLVTGTGALGAHVAPLAGRAEHLVLSRRGEDAPGAGALRAELAEALGA 2816
 QY 1135 RYV-----SDRSDFVADVAVAMTGG-----EGVDV-- 1161

```

Db      2817 RUTLAACDVADRASVAALLARLDAOGDPRAVAHVA--GGVVAQAALASTDLOCVQAVVA 2874
Qy      1162 -----INSLGELIDKSPNLRSHGRVEIGKRCQYADNOLGRLPRLNLSFSLVDLR 1214
Db      2875 KYRGALHLLHELLGDRLELDAFVLPAASGAGVWGGQGAAYAGNAFLD-----ALAEVR 2926
Qy      1215 -----GMM-----LERPARVALLEELLGLIAGVF-----TP 1242
Db      2927 RAAGLTATSTIAMGAMAGGGGLADHADADAEVELKRGGLVPMAPALIAALACALDHGETA 2986
Qy      1243 PPIATLPARVADAFRMAQAQCHIKLVLTLDPEVOIRIPT-HAGAGPSTGDRDLIDL 1301
Db      2987 VVYASVDWARFAPAFASARPR-----PLLDLPAGQAIEAPSAIPAGADGASADRALVEAL 3042
Qy      1302 ASAAPARAALAEAFRTQVSQVLRTPETKVGAEAL-----PTRLGMSIAMEVLEKRIE 1356
Db      3043 RPLSEADRAHVLVELVVAETAAVLGHAD-----ASALDPHKGFPDLGIDSLVAVELLKUR 3098
Qy      1357 ASIKLSTSTFLSTSPNIALAQNLDALATATSLERVAENLRA 1401
Db      3099 ARTVELPATVTFDHPSPHRLATYREALPLDAPLAPADATTAKS 3143

```

RESULT 10

```

OQ3NX8 PRELIMINARY; PRT; 5644 AA.
AC OQ3NX8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Amphib.
GN AMPH.
OS Streptomyces nodosus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=40318;
RN [1]
RP SEQUENCE FROM N.A.
RA Caffrey P., Lynch S.V., Flood E.M., Finnian S.M., O'Leary M.;
RT "The amphotericin biosynthetic gene cluster from Streptomycetes
   nodosus."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases..
DR EMBL: AF57202; AAK73502.1; -.
DR InterPro: IPR001227; AC transferase.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR004410; fadD.
DR InterPro: IPR000794; ketoacyl-synt.
DR InterPro: IPR006162; Pantane attach.
DR InterPro: IPR006163; Pp bind.
DR Pfam: PF00698; Acyl_transf. 3.
DR Pfam: PF00107; adh_zinc; 1.
DR Pfam: PF00109; ketoacyl-synt. 3.
DR Pfam: PF02801; ketoacyl-synt_C; 3.
DR Pfam: PF00550; Pp-binding; 3.
DR TRGFAMs: TRGR00128; fadD; 3.
DR PROSITE: PSS0075; ACP DOMAIN; 3.
DR PROSITE: PSS00606; B_KETOACYL_SYNTHASE; 3.
DR PROSITE: PSS00012; PHOSPHOPANTETHEINE; 3.
DR Phosphotransferase; Transferase.
KW SEQUENCE 5644 AA; 585775 MW; CFEA35E3BA0D2658 CRC64;
SQ

```

```

Query Match      32.7%; Score 2357; DB 2; Length 5644;
Best Local Similarity 33.0%; Pred. No. 2.3e-106;
Matches 632; Conservative 216; Mismatches 509; Indels 558; Gaps 39;

```

```

Qy      11 EDPAIYAGACRLPGGVIDYSGFTLLDGSRDTGVRVPAER-WDAAMPDPDDAPGKTP 69
Db      1797 DDVVIVGMGCRPGGSGPEDLMDLLASGDAITGPPADREMESSLV--AGEAGVGS- 1853
Qy      70 VTRASFSLDVACPDASFSGISPREALRMDPAHRLLEVCWEALENAAIAPSAIVGTGTG 129
Db      1854 -ACGFLSDIAGFDADFFGISPREALAMDPOQRITLLEVTWAIERAGVDPTALRGSRGTG 1912

```

```

Qy      130 FIGIGPSEYEALPOATASAEIDAHGSGCTMPSVGAGRISYALGRCPCVAVDTAVSSSL 189
Db      1913 FMVGQGVDTSSLV--MGSRDDVAGHATAGTAVSVSGSLSTYGLGEPALSTVDTCSSSL 1970
Qy      190 VAVHLACSLRSGECSTALAGVSLMSPSTLWLSKTRALARDCRCIAFSADQFGRG 249
Db      1971 VSLHLAAQRLRGECSSMLVGGVTWTTTPANFAGSRMGLQADRCRCAFSADAGTGW 2030
Qy      250 ECGAVVYLKRLSGAADDRILAVIRGSAIINHDSGSLTVPNGSSOEVLKRALADGC 309
Db      2031 EGAALVVERLSDARAGRIYAVVGVAVNDGASNGITAPNGSQQRVIRIQALANAL 2090
Qy      310 AASSGVYEAHGTGTLGDPRIEQALNAVYGGIRVATPPLIGSVKTNLGHREYASGTG 369
Db      2091 REGVDVAERHGTGTPPGDPIEQALLATGSDRPPQPLIGSVKSNIGHTQAAAGVAG 2150
Qy      370 LKRVVLSLQHGQIPALHLAQAALNPRI SWGDLRLTYTRATPWPDMNTPPRACVSSFGSG 429
Db      2151 LVKMMAMRNGVLPRLTHITTEPSTHYDMSLGAVVQVLTETETAMPETGRVBARAGVSSFGISG 2210
Qy      430 TNAHVYLEAPAACTPPAPERPAELV---LSARTASALDAQARLDHLETYSQCG 486
Db      2211 TNAHVILEGAPDEPVAPADRPVPGAVAMPVSASSEGALDQAEERLESADALPAL--- 2267
Qy      487 DVAFSLATRSAMEHRLAVALAATSRGRLRALDAAAGQTSPGAVERSIADSRGKLAFLT 546
Db      2268 DTAIVTLATRADPEIR-----AVLLADG-TLTVAGVAPRHPS--AFIFS 2311
Qy      547 GGAQOTLGMGRGLYDVMSAFREAFDLCYRLFNOELDRPLEVWMAEPASVDALLDQTF 606
Db      2312 GGAQGLGMGRGLHARFPVFAEAFDSVPLTLESELDTSVREVMW-----GTDEGALNATAF 2367
Qy      607 TOPALFTFEXYALALMRSGVEPELVAGHSIGELVAAACVAGVSLDEAVFLVAARGRLMQ 666
Db      2368 TOPALFAVEVALYRLVESGVTPDFVAGHSVEIAAHAGVFSLEDACRLVAARGRLMQ 2427
Qy      667 ALPAGAMVSIAPREADVAAVAAPHAASVIAVNAVDQVLAGAQOPHAIAAAMADARG 726
Db      2428 ALPAGAMVAVERTEDEVLPUL---TEGVAVAANVGPISVVSGBEQALTAVAEQALAQ 2484
Qy      727 ARTKALVSHAFHSPIMAELEAFGRVAVSVSYRRPSIYVNSLSKACTDEV-SSPGW 785
Db      2485 RRTSRLRVSHAFHSPIMAELEAFGRVAVSVSYRRPSIYVNSLSKACTDEV-SSPGW 2544
Qy      786 VHAAREVVPFAQGVKALHAAGATFVEVGPKSTLGLVPCMPDAPRALASSRAGRDP 845
Db      2545 VHAAREVVPFAQGVKALHAAGATFVEVGPKSTLGLVPCMPDAPRALASSRAGRDP 2604
Qy      846 ATVLEALGLIMAVGGLVSW-AGLFPSSGGRVPLPTYPMQRERYW----- 888
Db      2605 VSVLSGVARLVYAGVDVDSAPLSGAGRIADVPTTAPFHERRYWPAAPALDATGLGLA 2664
Qy      889 ----- 888
Db      2665 SADHPILGAAMSVAGSDELLTGLSLAATHPWLDHVGMIFFPOTGLELAVRAADQA 2724
Qy      889 ----- 888
Db      2725 GCDRVEELMIAPLVLPATGAVOVQISVGADEBGSRELPFTTREGEDDAEKTQATQR 2784
Qy      889 -----IDTKA-----DD----- 895
Db      2785 IGSGEVIDFDATWPPRAEALDIDGMFERAAODGLEGVFPRGLRAVMRQDDTYVAEV 2844
Qy      896 -----AANGDRRA----- 903
Db      2845 ALPESVEDADAFGLHALFDALHSTVFLSAEDGTRSLPFAWEGVSLHADASTLRVRI 2904
Qy      904 -----PG----- 905
Db      2905 ASGGEVDVQIAAVDPGGQPVSVESLTLRAAGGDAABRRDDNSLLRVDWTARTLGAAP 2964

```



```

Db 678 FSLIEDACTVAAAGAGMOLPRGAMVAATEDEVSLPLTLD---GVAIAINGPTSLV 734
Qy 709 AGAGGVVHAIIAAAMARGARTKALHSHAFSLMPLMEAERVAESVSRPSVIVS 768
Db 735 SGGETATLVAAALAOGRRTTRLRVSHAFSLMPLMEAERVAESVSRPSVIVS 794
Qy 769 NLSGKACTDE-VSSPGVWVHAREVVRFADVKALHAAGAGTVEVGPSTLLGLVPACM 827
Db 795 NLTGAVADGTLCTADYWRVHVRVFAVRFDGIRALDAGVGLLELGPDTLAAALQOSA 854
Qy 828 PDA-RPALLASSRAGDEPATVLEALGIMAVGLVSWAGLP-SGGRVRLPYTPWQR 884
Db 855 PDVSVFVL---RKORDEEPAVAALARLHTAGVPVDWTAFAAGCAHRTDLPYAFQY 910
Qy 885 ERYWID-----TKADDAARD-----900
Db 911 ERYWPKATYRPADATGTLTAADHPLLGAAMSVASDELLTGLSLATHPLMADHVVG 970
Qy 901 -----RRAPGAGHDEVE-----913
Db 971 MVEFPGTGLELAVRAADQVCDRVEELMAAPLILPATGVQMOLAVGAADDGGRDLR 1030
Qy 914 -----G 914
Db 1031 FTRPGDDPDAAWAQHATGRITEGERVLTDTTTPPRDAEPVLDGLYDRYANGLDYG 1090
Qy 915 GAARG-----GDRS---924
Db 1091 PVERGLRAVWRDTEIYAEVALPEGTADADAFGLHPALFDVAIHSTLPASADGDDSLR 1150
Qy 925 -----ARLDH 929
Db 1151 FANNVSLHAAGADALRVRTITSGCPDAVEITAVDPQGRPVSVESLTLRAAGPDAGADH 1210
Qy 930 -----PPESGRREKTEAAG- 944
Db 1211 RADAGSLFRMDWTPRTVHAAPATPATWALGTDPLGLTEALTAAGPDTVTGLRDGDALGE 1270
Qy 945 -----DRPF-----948
Db 1271 LTAAGDRPVPDVAVPLRGATDHGPAGADHLLTFTVALLQEWLAEEFPAASRLLLTRGA 1330
Qy 949 -----RLIEDP---955
Db 1331 VADGERGPLDLLAAPWGLVRSAGSEMRGRLLVDDDTESAQPLPLPALLDADBPQA 1390
Qy 956 -----GVLDHL-VLRTERRAPGLG-EVE 977
Db 1391 VREGTIVRGRLARLDSGRGLVPPPTPMRLGSRAGSLDGLALLPPEARRPLTGHEVR 1450
Qy 978 IAVDAAGLSFNDVQALGMV/PDOLPGKPNPPLLGGECAGRIYAVGVGVGVGVGVIA 1037
Db 1451 VGRAGAGLNFRLDNLNAGMTPGDAG-----LFGSAGGVVEVGEVGLAPGDVWG 1503
Qy 1038 LSAGAFATHTTTSAAVLPRPOLSAIEAAMPAVATLAWAYLADRIARLOPGEKVLTHAA 1097
Db 1504 MLGCGFEPRLIADARLLTLPVADMSWETGASVPLVFLTAYALKEGLGAGVIAVHAG 1563
Qy 1098 TGGVGLAAYVMAOHVGAEVHATAGTPEKRAYLESYGR--YVSDSRSDRVADVANTGC 1155
Db 1564 AGGGAIAIOIARVGAIEVPAIA-SEKMDVLSLGVADHDIASRSLTDEEAPFAEVAGD 1622
Qy 1156 EGVADVNLNLSGELIDSFNLSHGRFVELGKDCVADNOLGLRPLRLNLSFVLVDRG 1215
Db 1623 RGLDVVNLNLSGELIDSFNLSHGRFVELGKDCVADNOLGLRPLRLNLSFVLVDRG 1676
Qy 1216 MLEERPARVALLLEELGLIAAGVFTPPPIATLPIARVADAFRSMAQOHLGKLVLTG- 1274
Db 1677 MV--DEPHIQMLDLVELFDRGALALPVRSMVVRAGAEFRMSLAQHIGKIVLTPQ 1734
Qy 1275 --DPEVOIRIPTHAG--AG-----DSTGDRDIL--DRLASAAPARA----- 1310
Db 1735 PLDPDGTVLTLTGCGGLAGLLARLVTENHGBRHLLLAGRGRPDAPGAALHAELTALGAE 1794

```

```

Qy 1311 ---AALEAFRTQVSOVLRTPEIK-----VGAELFTRLGMSLMAVELRNRIE 1356
Db 1795 VTVAAACDVADRITALAALATVPAEHPLTAVHTAGVLDGDTLTALNPDLATV-LRPKD 1853
Qy 1357 AS-----LKLKSTFTSTST-----PNIALLA--ONLIDALAT---ALSL 1391
Db 1854 AAWHLHDLTRHLDLAFLVLYSSTAGVWGPGQANVAAGNTFILDALAAHRHALGL 1907

RESULT 12
Q914W3
ID Q914W3 PRELIMINARY; PRT; 11096 AA.
AC Q914W3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE NYSC.
GN NYSC.
OS Streptomyces noursei.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1971;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 11455;
RX MEDLINE=20334850; PubMed=10873841;
RA Brautaset T., Sekurova O.N., Sletta H., Ellingsen T.E., Strom A.R.,
RA Valla S., Zotchev S.B.;
RT "Biosynthesis of the polyene antifungal antibiotic nystatin in
RT Streptomyces noursei ATCC 11455: analysis of the gene cluster and
RT deduction of the biosynthetic pathway.";
RL Chem. Biol. 7:395-403(2000).
DR EMBL; AF263912; AAF71776.1; -.
DR HSSP; P25715; IMLA.
DR InterPro; IPR001227; Ac_transferase.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR004410; Fabd.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR006162; pantane_attach.
DR InterPro; IPR006163; pp_bind.
DR Pfam; PF00698; Acyl_transf. 6.
DR Pfam; PF00107; adh_zinc1. 1.
DR Pfam; PF00109; ketoacyl-synt. 6.
DR Pfam; PF02801; ketoacyl-synt_C. 6.
DR Pfam; PF00550; pp-binding. 6.
DR TIGRfam; TIGR00128; fabd. 6.
DR PROSITE; PS50075; ACP_DOMAIN. 6.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE. 6.
DR PROSITE; PS00012; PHOSPHOPANTETHEINE. 5.
KW Phosphopantetheine; Transferase.
SQ SEQUENCE 11096 AA; 1150415 MW; 776CAEAFCAE551DD CRC64;

Query Match 32.3%; Score 2331.5; DB 2; Length 11096;
Best local similarity 32.9%; Pred. No. 1e-104;
Matches 629; Conservative 182; Mismatches 519; Indels 583; Gaps 34;

Qy 10 AEDPIAVASGRLPGGVLDLGSFVTLGSRDTPVGRVPAER-WDAAWFDPDPDAPGKT 68
Db 3499 ADDPIVVGAGACFPFGGVTSPEDLMLRVDDGTDATITFTPNKMGDNDLXDPPEHFGTS 3558
Qy 69 PVTASFLSDVACFDASFGISPREALRMDPAHRLLEVCWEALENAAIAPSLVGTETG 128
Db 3559 YTRSGGFTHHAGFDFAPFGMSRRLATDSQGRLLLESWEAIEAAGIDPLTRGSATG 3618
Qy 129 VFTIGIPSEYEALPQATASAEIDAHGIGTSPSVAGAGRIAYALGRGVAVDTAYSSS 188
Db 3619 VFAGWVSYDGSTL---GGEKEFGQOGGSAGSVASGRVYALGEGPAVTVDTACSSS 3674
Qy 189 LVAVHLACGSTRSGESTLAGVSLMSPSTLWLSKTRALARDGRCAFSAAADQFGR 248
Db 3675 LVALHWAQALRLRAGESLALAGGVTVWSTPSTFVFSRQGLAPGRSRAFAEADQVGM 3734

```


QY 1292 TGDRLDRLASAPARAALAEFLRTQVSYLRPEIKVGAALFTRLGMDSLMAVEL 1351
 DB 5204 TGTGTGLGGL-----VARRHMERGVRRLL-----LTSRGGLDAAGAREL 5243
 QY 1352 RNRIEASLKLKLTSTPLTSPNIALLAQNLDLALATLSERVAEN-LE-----AGVON 1405
 DB 5244 VAELE-----NLGAEVSVAAACDVADBDVDTLLAGIPAEHPLRAVHTAGVLD 5291
 QY 1406 DFV 1408
 DB 5292 DGV 5294

RESULT 15
 OS3490 PRELIMINARY; PRT; 4151 AA.
 ID OS3490
 AC 053490;
 DT 01-JUN-1998 (TREMREL. 06, Created)
 DT 01-JUN-1998 (TREMREL. 06, Last sequence update)
 DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
 DE Polyketide synthase.
 GN PKS12 OR RV2048C OR MT2108 OR MTV018.35C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_TaxID=1773;
 RX MEDLINE=98295987; PubMed=9634230;
 RC STRAIN=H37RV;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala P.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulterson J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., Deboy R., Dodson R., Gwim M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL021899; CAI17262.1; -
 DR EMBL; AE007061; AAK46387.1; -
 DR HSSP; P25715; IMLA.
 DR TIGR; MT2108; -
 DR Tuberculist; RV2048C; -
 DR InterPro; IPR001227; Ac_transferase.
 DR InterPro; IPR002085; Adh_zn_family.
 DR InterPro; IPR007994; ketoacyl-synt.
 DR InterPro; IPR006162; Pantane_attach.
 DR InterPro; IPR006163; Pp_bind.
 DR Pfam; PF00698; Acyl_transf_2.
 DR Pfam; PF00107; adh_zinc_2.
 DR Pfam; PF00109; ketoacyl-synt_2.
 DR Pfam; PF02801; ketoacyl-synt_C_2.
 DR Pfam; PF00550; pp-binding_2.
 DR PROSITE; PS00075; ACP_DOMAIN; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS00012; PHOSPHOPANTHEINE; 2.
 DR Phosphopantetheine; transferase; Complete proteome.
 FT CONFLICT 2147 2147 H -> Q (IN REF. 2).
 FT CONFLICT 2260 2261 DI -> TV (IN REF. 2).

FT CONFLICT 2268 2268 W -> H (IN REF. 2).
 FT CONFLICT 2272 2272 S -> A (IN REF. 2).
 FT CONFLICT 2279 2280 AF -> PY (IN REF. 2).
 FT CONFLICT 2282 2283 AA -> GR (IN REF. 2).
 FT CONFLICT 2287 2287 T -> V (IN REF. 2).
 FT CONFLICT 2289 2289 F -> W (IN REF. 2).
 FT CONFLICT 3004 3004 S -> L (IN REF. 2).
 FT CONFLICT 3649 3649 P -> A (IN REF. 2).
 SQ SEQUENCE 4151 AA; 431577 MW; 536F644F111D7A5D CRC64;

Query Match 32.2%; Score 2321; DB 16; Length 4151;
 Best Local Similarity 33.2%; Pred. No. 8.7e-105;
 Matches 613; Conservative 203; Mismatches 535; Indels 498; Gaps 31;

QY 6 IERAAEDPRAIVAGASGRLPGVLDLGGFWLLGSGSDPTVGRVPAER-WDAAMFDDPPA 64
 DB 2051 VRTTSEDPPIAVGMACRYPGVNSPDDMDMLQGNVDLSEFPADGMDLGYNDPPA 2110
 QY 65 PKTPTPTASFLSDVACPDASFSGISPREALRMDPAHRLLEVCWALENAAIAPSAVG 124
 DB 2111 AGACTRTGTFDVGDFDPAFFGVGSEALAMDPOHRLLELSWEALERAGIDPTGLRG 2170
 QY 125 TETGVFTIGPSEYEALPQATASAEIDAHGGLGTPSYGAGRISTALGRPCAVDTA 184
 DB 2171 SATGVFAGVMTQGY-----GMFAAEPEVEGFRULGTQLSVASGVAVYGLLEGPVSVDTA 2225
 QY 185 YSSSIYAVHVAOCOSLRSGECSTALAGVSLMSPSTLWLSKTRALARDGRCAFAEAD 244
 DB 2226 CSSSLVALHMAVAGSLSGECDLALAGVTVNATPDIFVFSKRGISPDGRCAFAEAD 2285
 QY 245 GFGRGEGCAVVLKRLSGARADGDRILAVIRGSAINHDGASGLTVPNGSSOEIVLKRAL 304
 DB 2286 GTGFSGGGMVLVQRISDARLGHPLAVVAGSVAVNGDASNGLTAPNGSQORVVRAL 2345
 QY 305 ADNGCAASSVGYEAHGTGTTGDPLEIOTALNAVGLG-DAVATPLLISVKTNLGHPY 363
 DB 2346 ANAGLSAAEVDVEGHTGTTGDDPIEAQALLATYQODGDEPEDEPLMLSVXNMGHTQA 2405
 QY 364 ASGITLLVVLVLSLQHQJPAHLHAQALNPRIKSGDLRLTVRARPTPMDMTPRAGVS 423
 DB 2406 AAGVAVITKVLAMRELLPATLHDVPSPHDMSGAVELTAPVWPAGARTTRAGVS 2465
 QY 424 SFGMSGTNAHVLEAPPACTPPAPERPAELVLSARTASALDAQAARLRLHLETPSQ 483
 DB 2466 SFGISGTNHVITIEAVPVPRRAGWAGVVPVNVSAKSESALRGQAARLAAVVRDGG 2525
 QY 484 CLGDVAFLATRTSMEHRLAAVATISREGRLAALDAAGQCTSPGAVRSTADSSRKLAF 543
 DB 2526 DVADVGMSLA-GRSVEHRAVAVVGGDRDLGLGLDLADDLQGSVVRGTA-TAAKTYF 2583
 QY 544 LFTGGAQNTLGMGRGLYDVWSAFREAFDLCVRLFNQELDRPLREVMMAEPASVDALDQ 603
 DB 2584 VEPGGSQMLGMGMGLHAGYPVFAEFNTVYVGEIDHILRLPLREVMMGH-----DENLNS 2639
 QY 604 TAFTOPALFFFEYALALMRSMGVBEPELVAGHSIGELVAAVCAVGSLEDAVFLVARGR 663
 DB 2640 TEFAQPALFAVEVALRLLGSGWVRDPFVWGHSIGELSAHVAGVLSLEVAALVLAARGR 2699
 QY 664 LMQALPAGAMNISIEPEADVAAVAPHAASVIAAVNAPDOVVIAGACQPVHAIATAAA 723
 DB 2700 LMQALPAGAMVAVQAEEVEPRL--SAEVDIAVNGVASLIVISGAQNAVAVAVDQLR 2756
 QY 724 ARGARTKALVSHAFFSPLMAPLEAAGRAVESVSRRPSIVVSLSGKACTDEVSSPG 783
 DB 2757 ADGRRVHQLAVSHAFFSPLMDPVIDEPAAVAGIAGRPIIGVIVSVTGLADDDGSA 2816
 QY 784 YVVRHAREVVRFAAGVYALAAAGAGTVEVGPSTLLGLVPACMPDAPARALLASSBARD 843
 DB 2817 YMRHRHQAARFAVDSRFAQAAGSRFLVGPSSGLVASIEESLPDVAATVTMSALRKDRP 2876
 QY 844 EPATVLEALGGLMAVGSLVSWAGLPPSGGRVPLPTYPWQREKYYI--DTKADAA---- 897
 DB 2877 EPATLINAIVAGFVTGMDIDMRAVV-GEAQFVELPTYAFQRRRFLSGDVAADAAGLGL 2935


```
QY 898 -----RGDRRAPAGH-----DE 910
Db 2936 AASEHLLGAVIDLPAAGGVLTGRLSPSVQGLADHSAVAGVTIFPGAGFVELATRAGDE 2995
QY 911 V-----EG-----GAV 917
Db 2996 VCGGVVESTLAFLVLPASGSVAQVVVNGPDESGRGVSYSRGDVGITGWLHAEGL 3055
QY 918 RGS----- 920
Db 3056 RAGSAEPTLADLAMPAPGAVPEVEADGYOQLAERGVCYGPFRGLTAMMRGDEVEFAEVA 3115
QY 921 ----- 920
Db 3116 LPADAGSVTGFVHPVLLDPAHLHAVLSAESERGGQSVLPFSWQVSLHAAGASAVR 3175
QY 921 -----DR----- 922
Db 3176 ARIAPVPSAVSIELADGLPLVLSVASMILARPVTDOQLAAVSSSGPDLFEVTSPOP 3235
QY 923 ----- 922
Db 3236 SAAVEPLPVCANGTEDSAAVFESVPLAGDVVAGVYAATSSVLDVLSMLTRDGAGLV 3295
QY 923 -----RSARLDHP-----P 931
Db 3296 VMTRGVALPGEVDVTLDAAGAVWGLVRSAGTEHPRIVLVDSAPLDDSALAAVTTTGP 3355
QY 932 PESGRREKTEAA-----GRRPRLIEPQVLDHLVLR-VTERRAP-GLG 974
Db 3356 QVLMRGEVYTARVHSGRAVGLLVPSPDRPMLAMSTAGTFENLRLIPLADAPLPG 3415
QY 975 EVELINDAGLSFENDQALGFWPDDLPGKPNPRLLGECACRIYAVGEGVNGLVGQP 1034
Db 3416 QVRVAASAIANFRDWMIALGLPD-----PDAMGVEACGVLETSLNKGSEFAVGDR 3468
QY 1035 VIALSAGAFATHTTSAALVLRPPALSAIEAAMPAYILTAMAYALDRIRARLOPGERVLI 1094
Db 3469 VMGLPREGTGYASTQRLLVKVPAGWSHTAAATTSVPATAHYALVDLAAASGGRVLI 3528
QY 1095 HAATGGVGLAAVQMAQHVGAEVHATAGTEPKAYLESIGV--RYVSDSRSDRFVADVRAM 1152
Db 3529 HAGTGGVGAAYQVLAHHLGLEVFATA-SKGKMDTLRAMGFDDDHISRSLEFEDKFRAA 3587
QY 1153 TGGEGVDVYVNLISGELLIDKSFNLSHGRFVELGKDCYADNQLGR-PFLRNLSFSLV 1211
Db 3588 TGGRGFDVLDLSLAGEFVDASLRVAPGVFLEMGTDIRDPGVIAQYPGVRYRAFDLF 3647
QY 1212 DLRGWMLERPARVRALLLEELGLIAGVFTPPRIATLPIARVADAFRSMQAQHLGKLV- 1270
Db 3648 E-----PGPDRIAQILAEIATLFGDGVRLPLVPTTFDVRCAPALRLYLSQARHTGKVM 3701
QY 1271 -----LTTGDPVEQIRIPTHAGA-----GP-STGDRDLRLLAGA 1304
Db 3702 LMPGSWAAGTVLITGGTMAGSVAARHVVARHGVRLVLSRRGPDAPGAELVAMELAA 3761
QY 1305 AAPAARAAALEFLRTQVSOVLRTPEIK-----VGA--EALFTRLGMDSLMAVELRN 1353
Db 3762 GAQOVVACDADRALAKAVIADIPVQHPLSGVITHAGALDADAVVMSLTPDRVDVV-LRS 3820
QY 1354 RIEASLKLSTFTLSTSPNIAL-----LAQNLDPALA 1386
Db 3821 KYDAAWHHLFELTRDLVSAFVMESSMAGLVGSSGQANYAAANSFLDALA 3869
```

Search completed: October 2, 2003, 17:31:44
Job time : 167.882 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 26.2576 Seconds
(without alignments)
2544.975 Million cell updates/sec

Title: US-10-014-717-2
Perfect score: 7210
Sequence: 1 VADRIERAEADPIAIVGAS.....GVQNDPVSSGADQDWETIAL 1421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2107.5	29.2	3567	1 ERY2_SACER	Q03132 saccharopol
2	1925.5	26.7	3491	1 ERY1_SACER	Q03131 saccharopol
3	1889	26.2	3519	1 OLS5_STRAT	Q07017 streptomyc
4	1875	26.0	3172	1 ERY3_SACER	Q03133 saccharopol
5	1852.5	25.7	2110	1 MCAS_MYCBO	Q02251 mycobacteri
6	1592.5	22.1	1538	1 PRSB_WYCTU	Q10578 mycobacteri
7	1583	22.0	1876	1 PSA_WYCTU	Q10577 mycobacteri
8	1180	16.4	1986	1 WA_EMBE1	Q03149 emericella
9	1177	17.4	1774	1 MSAS_PENPA	P22367 penicillium
10	1173.5	16.3	2109	1 PKSI_ASPPA	Q12053 aspergillus
11	1100	15.3	2505	1 PAS_FAP	P12785 rattus norv
12	1083	15.0	2181	1 STCA_EMBE1	Q12397 emericella
13	1039	14.4	2504	1 PAS_HUMAN	P49327 homo sapien
14	1018	14.1	4447	1 PKSR_BACSU	P40803 bacillus su
15	997.5	13.8	2511	1 PAS_CHICK	P12776 gallus gall
16	940	13.0	4427	1 PKSI_BACSU	Q05470 bacillus su
17	882.5	12.2	4273	1 PKSM_BACSU	P40872 bacillus su
18	612.5	8.5	352	1 PAS_ANGAN	P36189 anser anser
19	392.5	5.4	317	1 PABD_BACSU	P71019 bacillus su
20	363	5.0	401	1 NOB4_RHIME	P06331 rhizobium m
21	354.5	4.9	402	1 NODE_RHIME	P06330 rhizobium m
22	344	4.8	308	1 PABD_SALTY	O85140 salmonella
23	336	4.7	403	1 NODE_RHIVV	P04683 rhizobium l
24	332	4.6	308	1 PABD_ECOLI	P25715 escherichia
25	329.5	4.6	467	1 KASA_STRCO	Q02059 streptomyc
26	328	4.5	312	1 PABD_HAEIN	P43712 haemophilus
27	325.5	4.5	402	1 NODE_RHIS3	P07231 rhizobium s
28	324.5	4.5	401	1 NODE_RHILT	P04684 rhizobium l
29	321.5	4.5	426	1 KASI_STRGA	P16538 streptomyc
30	316.5	4.4	300	1 VARI_HUMAN	Q09536 homo sapien
31	315	4.4	410	1 PABD_ECOLI	P39635 escherichia
32	308	4.3	414	1 PABD_VIBHA	P55338 vibrio harv
33	304.5	4.2	293	1 PABD_SYNY3	P73242 synecchocyst

34	303.5	4.2	413	1 PABF_VIBCH	Q0949 vibrio chol
35	303	4.2	838	1 PAS_MOUSE	P19096 mus musculu
36	302	4.2	420	1 KASI_STRCM	P41175 streptomyc
37	298.5	4.1	327	1 OOR_ECOLI	P28304 escherichia
38	298	4.1	421	1 KASI_STRVN	P16540 streptomyc
39	295.5	4.1	327	1 OOR_SALTY	P40783 salmonella
40	293	4.1	421	1 PABF_RHIME	P56902 rhizobium m
41	289.5	4.0	422	1 KASI_STRHA	Q05356 streptomyc
42	285.5	4.0	416	1 PABF_SYNY3	P73283 synecchocyst
43	284.5	3.9	423	1 KASI_STRCO	P23155 streptomyc
44	282	3.9	379	1 VARI_TORCA	P19333 torped cal
45	278	3.9	325	1 OOR_PSEAE	P43903 pseudomonas

ALIGNMENTS

```

RESULT 1
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC Q03132; 054096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteriales;
OC Pseudonocardiales; Pseudonocardiales; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis."
RL Science 252:675-679(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=NRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevil D.J., Cortes J., Haydock S.F., Leadley P.F.;
RT "6-deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme."
RL Eur. J. Biochem. 204:39-49(1992).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (PAS).
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M63677; AAA26494.1; -
DR EMBL; X62569; CAA44448.1; -
DR PIR; S23070; S23070.
DR InterPro; IPR001227; AC transferase.
DR InterPro; IPR002085; Adh zn family.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR006163; Pp_bind.
DR InterPro; IPR006162; Ppanthe_attach.
DR Pfam; PF00107; ADH_zinc_N; 1.
DR Pfam; PF00109; ketoacyl-synt; 2.
DR Pfam; PF02801; ketoacyl-synt; 2.
DR Pfam; PF00550; pp-binding; 2.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
DR PROSITE; PS00606; B KETOACYL SYNTHASE; 2.
DR TRANSFERASE; Acyltransferase; 2.
KM Phosphopantetheine; Multifunctional enzyme.
FT DOMAIN 1 1484 MODULE 3.
FT DOMAIN 1485 3567 MODULE 4.
FT DOMAIN 27 488 BETA-KETOACYL SYNTHASE 1.
FT DOMAIN 559 884 ACYLTRANSFERASE (AT) 1.
FT DOMAIN 1130 1301 BETA-KETOACYL REDUCTASE 1 (POSSIBLY NON-FUNCTIONAL).
FT DOMAIN 1397 1467 ACYL CARRIER (ACP) 1.
FT DOMAIN 1485 1943 BETA-KETOACYL SYNTHASE 2.
FT DOMAIN 2013 2336 ACYLTRANSFERASE (AT) 2.
FT DOMAIN 2383 3066 DEHYDRATASE/ENOYLREDUCTASE (DH/ER).
FT DOMAIN 3139 3322 BETA-KETOACYL REDUCTASE 2.
FT DOMAIN 3415 3485 ACYL CARRIER (ACP) 2.
FT ACT_SITE 202 202 THIOESTER BOND.
FT ACT_SITE 651 651 ACTI-ESTER INTERMEDIATE.
FT BINDING 1430 1430 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT_SITE 1661 1661 THIOESTER BOND.
FT NP_BIND 2961 2978 ACYL-ESTER INTERMEDIATE.
FT NP_BIND 3142 3157 NADP (ER).
FT BINDING 3448 3448 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 438 438 R -> A (IN REF. 2).
FT CONFLICT 480 480 T -> S (IN REF. 2).
FT CONFLICT 1241 1241 L -> F (IN REF. 2).
FT CONFLICT 2664 2664 G -> V (IN REF. 2).
SQ SEQUENCE 3567 AA; 374413 MM; E6284F4738AA0C CRC64;

Query Match 29.2%; Score 2107.5; DB 1; Length 3567;
Best Local Similarity 32.5%; Pred. No. 4.5e-103;
Matches 610; Conservative 179; Mismatches 530; Indels 559; Gaps 45;

QY 2 ADRIPIEAAE-DPAIVGASGRLPGGVIDLSGFYTL-EGSRDVGVRPWR-WDAAMF 58
DB 1478 AVRVGAADSEPAIVGIGCRPGGIGSPBQLMRVLAEGANLTLTG-PPADRGNDIGRLY 1536

QY 59 DPPDAGKPTVTRASFSLDVACPDASFPGISPREALRMDPAHRLLEVCMEALENAIA 118
DB 1537 HRPDNDGTSYVDKGFUTDAADDPFGFTGTPREALAMPQOKLMETMEANERAGID 1596

QY 119 PSAIVGTGTGFTGIGSEYEALPQATASAEIDAHGCGTGMPVGAGRISYALGRPC 178
DB 1597 PDALRGDTGVFVGMNGOSYQULL--AGEABRVGVGQLGNSASVLSGRILATYFGWEGPA 1654

QY 179 VAVDTAASSSLVAVHLACQSLRSGECTALAGCISLMSLSTLVMLSKTALARDGCKA 238
DB 1655 LTVDTACSSSLVGHLMQALRREGCSLALAGVTJWSDPYTFVDFSTQGLASDGRCKA 1714

QY 239 FSAEADFGGEGGCAVVVLKRLSGARADGRILAVINGSAINHDGASSGTLVPNGSOEI 298
DB 1715 FSAADGFALSEGVALVLELISRAKANGHVLAVLGSVAVNQGASNGLAAPNGPQER 1774

QY 299 VLKRALADAGCAASSVGVVEAHGTTLGDPFIEIQLANNAVYGLGRDVATPLLLGSVXTNL 358
DB 299 VLKRALADAGCAASSVGVVEAHGTTLGDPFIEIQLANNAVYGLGRDVATPLLLGSVXTNL 358

DB 1775 VIRQALAAAGVPAADVVEAHGTGTETELGDPPIEAGALITATYGQDRD--RPLRLGSVKTNI 1832
QY 359 GHPEVASGITGLKLVLSLQHGQIPAHHLAQALNPRLISWGDLRLTTRARTPEMDWTRP 418
DB 1833 GHTQALAAAGVATKVLARHGLRSLHADELSPITIDESGAVEYLREVEWPPAGERR 1892

QY 419 RAGVSSFGSGTNAHVLEEAR-----AATCTPAPERPAELLVLSARTASALDAQARL 473
DB 1893 RAGVSSFGVSGTNAHVLEEARPAEQEARTENGPLP-----FVLSGSEAVVAQAARAL 1946

QY 474 RDHLETYPQCCGDVAFSIATTRSMENHRLVAALTSREGRLAALDAAGQOTSPGAVRSI 533
DB 1947 AEHLRDTPELGLTDAAMTATGRARPDVRAAVLGDRAVCALDLAEGRSDADVAHV 2006

QY 534 ADSRQKALFLFGGAGQTLGKGRGLYDVMSAFREAFDLCVLLFNDELPRPREVMWAP 593
DB 2007 TSAFPR-KPVLVFPGQAGQVGNARDLBESSEVFAEBSMCAEALSHHTWKLLDVRRGG 2065

QY 594 ASYDALLDQTAFTQPAFTFEYALALWRSWGBELVAGHSIGELVACVAVGSLELD 653
DB 2066 GPDHERVD---VLQPVLSIMVSLAELWRAGVTTPAAVVGHSQGEIAAAHVAAGALSLBA 2122

QY 654 AVFLVARGRLMOALPAGAMVSIAPBADVAALVAPHAASVIAAVNAPDOVITAGAQ 713
DB 2123 AAKVVALRQVRLRELDQGGVSVGSRSDELTEVTLARMQGRVAVVAVNGPSTSVAGPTA 2182

QY 714 PYHAIAMAARGARTKALVSHAFSPIMAPMLAEFGRAVSVSRRPSIVLVSNLGK 773
DB 2183 ELDEFFAEADAEKMRRLAVRYASHPEVARLEDLAELGITTVAVRSVPLHSTVTE 2242

QY 774 ACTDESSBGVWRHAREVVRPADVKAALHAAGAGTFVEGPKSTLL-----GLV 823
DB 2243 VIDTSMADSYWYRNLRRLRFLFQAVRGLVEQCFDTEFVESHVPLVLMVEETAENAGAE 2302

QY 824 PACMPARALLASSAGDEPATVLEAL--GGLNAVGLSVWAGLF--SGGRARVPLPT 879
DB 2303 VTCVP-----TLREQSGPEFLRLNLRHNVHGVG-----ADLRPAVAGGRPAELPT 2349

QY 880 YPMQRERYV-IDTKDDAARGDRRA----- 903
DB 2350 YPEHQRFRPRPRPVDVSLGVGAEHPLLAADVDPHGCAVFTGRISTDEQPLAH 2409

QY 904 -----PG-----AGHD----- 909
DB 2410 VVGRTLVGSLVDLALAAGEDVGLPVLBEVLQRPVLVLAGALLRMSVGADESGR 2469

QY 910 ----- 909
DB 2470 TIDVHAABVDLADAOQSHATGTLAOGVAAGPRDTEQMPEDAVRIPLDHYGLAEQ 2529

QY 910 -----EV-----EEG-----GAV----- 917
DB 2530 GYEVGSFOALRAAMKDSVYAEVSIADDEGYAHFVLVDVAQTLSLGLNGERGCGK 2589

QY 918 -----RGDRRSARL----- 930
DB 2590 LPAAMTVTLHAGATSVRVATPAGADAMALRVTDPAGHLVATVDSLVRSTGEKKEOP 2649

QY 931 PPSGGR----- 947
DB 2650 EPRGGEGLHALDMGRLABEGSTGRVVAADSDDLVLSGEBEPDPAVLVRYEPEBDPR 2709

QY 948 ----- 947
DB 2710 AAARHGVLMALVRKMLEOEBELPGATLVATSGAVTSDDSVPERGAAMMGVRCQA 2769

QY 948 -----FRLEID-EPGVL-----DHLVLRV-----TERAPG 972
DB 2770 AESPDFFVLIDTPAEBGMLPAYPDNPQALARGDVFVPRLSPLAPALTLPACTQRLVPG 2829

QY 973 -----LGEVELAVDAAGLSFNDVOLALGMVDDDLGKKNPULLG 1012
DB 2830 DGAIDVAEPADVDQPLRAGEVVRVDVATGVNFRDVLALLGMYEQAD-----WG 2881

QY 1013 GECAGRIAVGEGVNGLVGQGVIALSAGAFATHTTSATLVLPFOALSAIEAAMPVA 1072
 DB 2882 TEAAGVAVAVGPDVDAFARGDVLGFGQAFPIANTDRLRLARVPDGSDDAAAVPIA 2941
 QY 1073 YITANVYALDRILARLOGEYVLIHAATGVGLAAVQAGVAGVATAGTPEKAVLESL 1132
 DB 2942 YTTAAHYALHDLGLAGGOSVLIIHAAAGVGMVAVALARRAGAEVLATAG-PAKHGTLRAL 3000
 QY 1133 GV--RYVSDRSRDRFPAVDVAVMTGEGVAVVNSLSEGLIDSEFNLRSHGFVEIGKD 1190
 DB 3001 GUDDEHILSSRETGFARKFRERTGGVAVVNSLTGELDSADLLAEDGVFVEGKTD 3060
 QY 1191 CYADNQLGRLPFLRNLSFSLVDLRGMMLERP-----RVALLLEELGLLAAGVF 1240
 DB 3061 -----LRLDG-----DFRG-----RYAPFDGEGAGDRLGELILEVVGILGAGSL 3100
 QY 1241 TPEPIATLPIAVADAFRSMAOQHLKVLVTLG---DPEVOIRLPTHAGAPSTGDRDL 1297
 DB 3101 DRLPYVAMELGSAPALQHMSSRGHVGLVLTOPAPVDDGTVL-----TGCTGT 3151
 QY 1298 LDRLASAPAAARAAALFRLTOVSVVLTPEIKVGAELFRLGMDSLMAVELNRRTA 1357
 DB 3152 LGRL-----LARHLVTEHGVRL-----LVSRRGADAGSDELRAELR- 3190
 QY 1358 SLKLTSTFTLSTPNIALLAONLIDALATL-----SLERVAEN-LR 1400
 DB 3191 DLGASAEIACDTARDALSA--LIDGLRPLTGVAHAGVADGLVTSIDEPAYEQVLR 3248
 QY 1401 AGVQNDVSSGADQDWEI 1418
 DB 3249 AKV-----DAAMNL 3257
 RESULT 2
 ERY1_SACER STANDARD; PRT; 3491 AA.
 ID ERY1_SACER
 AC Q03131;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-deoxyerythronolide B synthase 1) (DEBS 1).
 GN ERYA.
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).
 OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;
 OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
 OX NCBI_TaxID=1836;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91220065; PubMed=2024119;
 RA Donadio S., Scaver M.J., McAlpine J.B., Swanson S.J., Katz L.;
 RT "Modular organization of genes required for complex polyketide biosynthesis.";
 RT Science 252:675-679 (1991).
 RL [2]
 RN SEQUENCE OF 3474-3491 FROM N.A.
 RP MEDLINE=93231529; PubMed=8386127;
 RA Donadio S., Scaver M.J.;
 RT "IS1136, an insertion element in the erythromycin gene cluster of Saccharopolyspora erythraea.";
 RT Gene 126:147-151 (1993).
 RL -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-deoxyerythronolide B.
 CC -1- CORRECTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN BIOSYNTHESIS.
 CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THIS ERYA SHOWING 3 ORF CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,

CC RESPECTIVELY.
 CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
 CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
 CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
 CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
 CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
 CC OF THE FULL-LENGTH CHAIN.
 CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
 CC -1- SIMILARITY: Contains 3 acyl carrier domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M63676; AAA26493.2; -;
 CC EMBL; L07626; AAA26504.1; -;
 CC PIR; T43231; T43231.
 CC HSSP; P25715; IMEA.
 DR InterPro; IPR001227; Acyltransferase.
 DR InterPro; IPR002198; Acyl carrier.
 DR InterPro; IPR000794; Ketoacyl-synth.
 DR InterPro; IPR006163; Pp bind.
 DR InterPro; IPR006162; Ppantone attach.
 DR Pfam; PF00698; Acyl_transf; 3.
 DR Pfam; PF00106; adh_short; 1.
 DR Pfam; PF00109; ketoacyl-synth; 2.
 DR Pfam; PF02801; ketoacyl-synth; 2.
 DR Pfam; PF00550; pp-binding; 3.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR PROSITE; PS00075; ACP_DOMAIN; 3.
 KW Phosphopantetheine; Multifunctional enzyme.
 FT DOMAIN 1 1972
 FT DOMAIN 1 1979 3491
 FT DOMAIN 1 1 375
 FT DOMAIN 1 414 484
 FT DOMAIN 1 503 961
 FT DOMAIN 1 1030 1356
 FT DOMAIN 1 1611 1794
 FT DOMAIN 1 1888 1958
 FT DOMAIN 1 1979 2441
 FT DOMAIN 1 2507 2854
 FT DOMAIN 1 3055 3237
 FT DOMAIN 1 3334 3404
 FT ACT SITE 145 145
 FT ACT SITE 447 447
 FT BINDING 677 677
 FT ACT SITE 677 677
 FT ACT SITE 1128 1128
 FT NP_BIND 1614 1660
 FT BINDING 1921 1921
 FT ACT SITE 2148 2148
 FT ACT SITE 2598 2598
 FT NP_BIND 3058 3104
 FT BINDING 3367 3367
 SQ SEQUENCE 3491 AA; 365022 MW; 682BFC32C90FA8C4 CRC64;
 Query Match 26.7%; Score 1925.5; DB 1; Length 3491;
 Best Local Similarity 35.6%; Pred. No. 1.7e-93;
 Matches 547; Conservative 184; Mismatches 601; Indels 203; Gaps 39;
 QY 4 RPIEPAADPIAIVGASCRLLPGCVTLDSGFWTLGSRSDTYGRVAER-WDAAMPDPDP 62
 DB 497 RTNEAAPPEPAVAVVMAICRLPGCVSTPEEFELLSEGRDAVAGLPTDRGMDDSLFFHDP 556
 QY 63 DARGTPTVTRASFSDVACFDASFCGISPRELRMDPAHRLLEFCWELLENAAIAPSL 122
 DB 557 TRSGTAHORGGGFLTEATAPFAFPFGMSPREALAVDPORLMLLSWEVLERAGIPPTSL 616

KM Transferase; Acyltransferase; Antibiotic biosynthesis; NADP,
KW Phosphopantetheine; Multifunctional enzyme; Repeat.
FT DOMAIN 1 3519 ?
FT DOMAIN 2 501. ?
FT DOMAIN 3 569 890
FT DOMAIN 4 1200 1382
FT DOMAIN 5 1487 1561
FT DOMAIN 6 1686 2156
FT DOMAIN 7 2220 2541
FT DOMAIN 8 2856 3038
FT DOMAIN 9 3141 3215
FT DOMAIN 10 3270 3519
FT ACT SITE 210
FT ACT SITE 660 660
FT NP BIND 1203 1249
FT BINDING 1524 1524
FT ACT SITE 1859 1859
FT ACT SITE 2311 2311
FT NP BIND 2859 2905
FT BINDING 3178 3178
SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAE61F86 CRC64;

Query Match 26.2%; Score 1889; DB 1; Length 3519;
Best Local Similarity 34.0%; Pred. No. 1.4e-91;
Matches 541; Conservative 191; Mismatches 595; Indels 264; Gaps 44;

QY 9 AAEDIIATVAGSACRLPGVYIDLSGFMTLESGRDTGVGRVPAER-NDAAAMPDPDAPGK 67
DB 1664 ASREIATITANSRCRGGIDSPEDIMRFLABGRDVAAGIPEDRGMDLDLHYDPDENPQT 1743
QY 68 TPVTRASFSLDVACDFASFFGISPREALRMDPAHRLLEVCWALENNAIAPSAVLGTET 127
DB 1744 TVVREGARRYDAAGDAGFGISPREALMDPQGLLLETSELPERADIDITYTRGTAT 1803
QY 128 GVFTIGGSEY---EALPQATASAEIDAAGGLGTMPSYAGRISYALGLRCPVAVDTA 184
DB 1804 GIFTAGGQGYGDPDKRAPESSVAYIL-----TGTAASVLSGRISYTGLEGPATVDTA 1858
QY 185 YSSSLVAVHLACQSLRSRGCSTALAGVSLMSPSTLWLSKTRALADGRCARFASREAD 244
DB 1859 CSSLVLAHLAVOALRRCGCSLAIAGVAWVSTPAFEPFRQGMARDGRKATPAAD 1918
QY 245 GFGREGCAVVVVKELSGARADRIILAVIRGSAINHGSAGSLTPVNGSQEIVLKRAL 304
DB 1919 GNGWGEVSLILRLSLDABRLGHRVLAIVVRSANVQGSASGLAANGPQORVIRAL 1978
QY 305 ADAGCAASSVGVVEAHGTTTIGDPIETIOALNAVYGLGRDVAATPLLIGSVKTNLGHPEYA 364
DB 1979 ADAGLAAPADVVEAHGTTTIGDPIETIOALNAVYGLGRDVAATPLLIGSVKTNLGHPEYA 2038
QY 365 SGITGLKLVLSLQGOIPAHLLAQLNPRISSWGLRLTVTRAPRPW-PDMNTTPRRAIS 423
DB 2039 AGVAVGKVKVLLGKGVVPKTLHVDPSPHVMSAGAVELLTEERPMEPEARLERLAGIS 2098
QY 424 SEFGSGTAAHVVEEPAPATCTPPAPERPAELL-----VLSARTASALDAQAARLR 474
DB 2099 AFGVGTAAHVIVEEAPR-----EPEPEPTRVVAAGDLVVPVVGSGROVGLAREPAALLA 2154
QY 475 DHLETVPSQCLGDVAFSLATTSAMEHRLAVATSRREGRLAALDAQAQTSFGAVRSIA 534
DB 2155 AHVSTGTAGV-DVGMSTVATRSVFERRAVVVGTLDSWAGSLAFAAGVGVPGVSGVA 2213
QY 535 DSSRGKLAFLFTGGAOTLGMKRGLYDVWSAFREAFDLCVRLFNELDRPLREVMMAEPA 594
DB 2214 PAEGRRVVFVPPGQSQWGMAGLLDACPVFAEVAECAAALVDVTGMSLVEVQGR-- 2271
QY 595 SVDAALLDQTAFTQPALFTFEYALAAWRSMGVBEELVAGSHIGELVACVAGVSELEDA 654
DB 2272 --DAVTLGRVDVVGQALMAVWVSLARTWRYGVVEEAAVVGSHQGIATAACVAGGSLADG 2329
QY 655 VFLVAKRGLMQLPAGGAMVSIAPEDVAAVAAPHAASVIAAVNAPDOVVIAGAGOP 714

DB 2330 ARVVLSRAIARIAGGGWVSVSLPAGRVRTMLDTYGGRVVAAVNPGSSTVSGDVQA 2389
QY 715 VHAIAAAMAAAGATKALHVSFAFHSPLMAPMLEAFGRVAVSVYRSRSIVLNSGKA 774
DB 2390 LDELLAGERGCVARRRPVVIYASHSQMDQRLBELBALDITQDSSVFPEFTVTADW 2449
QY 775 CTDEVSSGCVRRHAREVVRPADGVKALHAAAGCTFVEVGRKSTLLGLVPACMP--DAR 832
DB 2450 LDTTALDAGVFTMLRETVRFOEAVEGLVAGQMGAFVFCSPHPVLVPGIEQTLDALDQNA 2509
QY 833 ALLASSRAGREPA--TVLEALGGLMAVGLVSN--GLPFGSGRVPLPTVWQER 886
DB 2510 AVLSGLR--REBGGDLRLTSLAEAFVQGVVDVWTHAFEGVTP--RTVDLPETYDFOR 2564
QY 887 YWIDTKADDAARG-DRAPGAGHDEVEEGAVR-----GGRRSARLDHPPESGRRKV 940
DB 2565 FWLDSPASSANGVDGEADAMIMDAVERDSVAABEIGIAEALHTVLPALSSRRRRV 2624
QY 941 E-----AAGRPFLRIDE-----PGVLDHLVLRVTERAPGLGE 975
DB 2625 EHRRLQDWRYVEMKFPALDDEVLGGMFLVPRGLADGVARVAAVTARG--GE 2680
QY 976 VEI-----AVDAAGLSFVDVQALGMP--DLPCKRNPPLLGECGA 1016
DB 2681 VSVVELDPTRPDRARAYEAVARGVS-----GVVSFLSMDRRHSEH-PVVPAGLAA 2731
QY 1017 GRIVA-----VGEGVN-GLVVGQPVIA--LSAGAFATHTVTS--ALVPRPO- 1059
DB 2732 SLVLAQALVDLGRVBEGRMLVTRDAVAGPSDAGVYIDVQAVQWFGFVGLHEPEL 2791
QY 1060 -----ALSAIEAA--MPVAYLTWALDRIAL- 1086
DB 2792 WGLIIDLPEAPPEPGSTCDHTYADLLATVVASAGREDDVAVRSGGV--VARLVAVVDG 2849
QY 1087 -----QCEERYLTHAATGCVGLAAVQAAQVGAEVH-----ATATPEKRAYLESL 1132
DB 2850 GGGGWRPRGTVLVYGLGLGCAHTARMLVCGGAD-HVTLVSRGGSGAPGADLVRELEGL 2908
QY 1133 GVRVYSDRSRPAVD--VRAMTG--GEGVDVVLNLS- 1166
DB 2909 GGARVSVRACD--VADRVALPALLSDLGEPTTAVFHAAGVQSTPLAETISQEAADVMA 2966
QY 1167 -----GELID-----KSFNLRSHGRFVELGRDCA- 1193
DB 2967 KVAGAVNLGELVDCGLEAFPLFSSNMGWMSGQAVYAAANAFDLALAVRRGGLPAT 3026
QY 1194 -----DNQLGRPPLNLSFSLVDLRGMLERPARVALLBELGLIAAGVTFPP 1244
DB 3027 SVANGMWAGEOMASVGGARELS-----RRGVRAMDEPRAVAVVADVGRGEAFV--A 3077
QY 1245 IATLPIARVADAFPSMAQOHLKVLVLTADPEVOIRI-----PTHAGAPSTGDRDL 1298
DB 3078 VADVMEFVTGFSAPRPLISDL-----PEVRTAARNQOEOQLHAPV-PEDRSACL 3130
QY 1299 DRLSAAAPARAALAEAFRLTQVSOVLRTPEIK-VGAELFTRLGMSLMAVELRNRIEA 1357
DB 3131 RRLMLSPAGREAEVLVLRTEAAVAVLGHSAQOVPRERAKELGFSLTAVQLRNRLAA 3190
QY 1358 SLKLKLTSTLSTPNTALLAQNLLDALATA 1388
DB 3191 ATGTRLPASAVFDHPHAALRLMLLAGWRHA 3221

RESULT 4
ID ERY3_SACER STANDARD; PRT; 3172 AA.
AC Q03133; Q54097; Q99270;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-deoxyerythronolide B synthase III) (DEBS 3).
GN ERYA.

OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=91043075; PubMed=2234082;
RA Cortes J., Haydock S.F., Roberts G.A., Beville D.J., Leadlay P.F.;
RT "An unusually large multifunctional polypeptide in the erythromycin-
producing polyketide synthase of Saccharopolyspora erythraea.";
RL Nature 348:176-178(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Stever W.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
biosynthesis.";
RL Science 252:675-679(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Beville D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
Cloning of the structural gene, sequence analysis and inferred domain
structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
deoxyerythronolide B.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
SYNTHASE PARTICIPATES IN ONE OF THE SIX PAS-LIKE ELONGATION STEPS
REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
BETA CARBON. AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X56107, CA39583.1, -;
DR EMBL, M63677, AAA26495.1, -;
DR EMBL, X62569, CAA44449.1, -;
DR PIR, S13595, S13595.
DR PIR, S22012, S22012.
DR HSSP, P00101, ICCH.
DR InterPro, IPR001227, Ac_transferase.
DR InterPro, IPR002198, ADH_short.
DR InterPro, IPR000794, ketoacyl-synt.
DR InterPro, IPR006163, Pp_bind.
DR InterPro, IPR006162, Pantne_attach.
DR InterPro, IPR001031, Thioesterase.
DR Pfam, PF00698, Acyl_transf. 2.
DR Pfam, PF00106, adh_short. 1.
DR Pfam, PF00109, ketoacyl-synt. 2.
DR Pfam, PF02801, ketoacyl-synt_C. 2.

DR Pfam, PF00550, Pp-binding. 2.
DR Pfam, PF00975, Thioesterase. 1.
DR PROSITE, PS00012, PHOSPHOPANTETHEINE. 2.
DR PROSITE, PS00606, B_KETOACYL_SYNTHASE. 2.
DR PROSITE, PS50075, ACP DOMAIN. 2.
DR Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KM Phosphopantetheine; Multifunctional enzyme.
FT DOMAIN 1 1484
FT DOMAIN 1485 3172
FT DOMAIN 317 484
FT DOMAIN 554 878
FT DOMAIN 1116 1298
FT DOMAIN 1394 1464
FT DOMAIN 1488 1954
FT DOMAIN 2021 2335
FT DOMAIN 2555 2735
FT DOMAIN 2821 2891
FT DOMAIN 2926 3172
FT ACT SITE 199 199
FT ACT SITE 643 643
FT NP BIND 1118 1164
FT BINDING 1427 1427
FT ACT SITE 1661 1661
FT ACT SITE 2112 2112
FT NP BIND 2557 2605
FT BINDING 2854 2854
FT CONFLICT 231 231
FT CONFLICT 240 240
FT CONFLICT 289 289
FT CONFLICT 493 493
FT CONFLICT 493 517
FT CONFLICT 510 510
FT CONFLICT 513 513
FT CONFLICT 525 525
FT CONFLICT 536 536
FT CONFLICT 547 551
FT CONFLICT 553 553
FT CONFLICT 673 673
FT CONFLICT 716 716
FT CONFLICT 734 736
FT CONFLICT 796 796
FT CONFLICT 896 896
FT CONFLICT 898 994
FT CONFLICT 1108 1116
FT CONFLICT 1124 1126
FT CONFLICT 1132 1132
FT CONFLICT 1192 1192
FT CONFLICT 1194 1194
FT CONFLICT 1277 1278
FT CONFLICT 1385 1390
FT CONFLICT 1485 1485
FT CONFLICT 1518 1518
FT CONFLICT 1601 1601
FT CONFLICT 1724 1725
FT CONFLICT 1732 1732
FT CONFLICT 1739 1743
FT CONFLICT 1762 1762
FT CONFLICT 2252 2252
FT CONFLICT 2275 2277
FT CONFLICT 2408 2408
FT CONFLICT 2420 2421
FT CONFLICT 2443 2444
FT CONFLICT 2596 2596
FT CONFLICT 2609 2609
FT CONFLICT 2715 2722
FT CONFLICT 2754 2754
SQ SEQUENCE 3172 AA; 331474 MW; DBBDS094E77DDSF CRC64;
Query Match 26.0%; DB 1; Length 3172;
Best Local Similarity 34.6%;
Matches 549; Conservative 184; Mismatches 602; Indels 254; Gaps 48;

QY 4 RPIERAEDPIAIVGASCRIPGVIDLSGFWTLLESGRDTGVRVPAERMDAAAFDPDP 63
 DB 31 REVEHAGEPIAIVGACRFPDVSPESEFVSGGDAIAEAPDR-----CM-EPDP 85
 QY 64 AOKPVTPTAASPLSDVACPDASFFGISPREALMDPAHLLLEVCHEALENAIAASAV 123
 DB 86 A-----RUGMIAAAGDPDAGFFGISPREALMDPOQRIMEISWEALERACHDVSUR 139
 QY 124 GTEYFVIGIGSEYEALPQATASAEIDAHGGLTGPSVAGRISYVALGRPCVAVNT 183
 DB 140 GSAITGVFTGVGTVDYPRPDE--APDEVLCYVGTGTASSVASGRVAVCIGLESPANTVOT 197
 QY 184 AVSSSLVAVHLACQSLRSEGCSTALAGVSLMSPSTLWLSKTRALARDGRKAFSA 243
 DB 198 ACSSGGLTALHAMESLRDECGALAGVTVMSPGAFTFRSQGLAADGRCKPFSKAA 257
 QY 244 DGRREGCAVYVVKLSGABADGRIILVIRGSAINHOGASGLTVPGSSQSEIYLKA 303
 DB 258 DGGIAGAGVAVLQRLSARRRGVLAVALAGSANDGASNGLTAPSGAQVIRRA 317
 QY 304 LADGCAASSVGYEAHGTGTTGDDPIEIOALNAVYGLGRDVATPLLIGSVKTNLGHPEX 363
 DB 318 LBNAGVRAGDVYVEAHGTGTRIGDPIEVHALLSTYGAERDPDDPLMIGSVSNIGHTQA 377
 QY 364 ASGITGLLVLSLOHGOIPAILHAQALNPRISWGDRLTVTRARTPMDMTPRRAGVS 423
 DB 378 AAGVAGVMKAVIALRGMEMPTLHPEPSPQIEMDGAIVSVQASWPAAGERPRRAGVS 437
 QY 424 SFGMGTNAHVLEENPAATCTPPAPERPELLVLSARTASALDADAARLDHLETPQC 483
 DB 438 SFGISGTNAHVLEENPEADEPPAPDSGPVLYLSGRDQANRAQAGRLADHLEPEPN 497
 QY 484 CIGDVAFSLATFSAMEHLAVATSR-----GLRALDAAQOTSPGAIVSIADSRCK 540
 DB 498 SLRDIGFTLATRASAMEHRAVVGDDDEALAGRAVADRIRADRTTQ-----GPNSPR 553
 QY 541 LAFLFTGGAGTGLMGRLGYDVMSAFREAFDLCVRLFNOLDRPREVN--WAEPAVDAA 599
 DB 554 VAMVFPQAGQWQGMARDLLRESQVADSIRDCERALAPHVMSLTDLSGARP----- 607
 QY 600 LDDQTFOTPALFTFYALALMRSWGVPELVAGHSIGELVACVAGPSLEDAVFLVA 659
 DB 608 -LDRVDVVPALFAVWVSLAALMRSHGVPEAAVVGSHOEIAAHVAGALTLEDAAKLV 666
 QY 660 ARGRLMQLPAGGAWVSIAPBEADVAAVAAPHAASVIAVANP--DQVVIAGAGCVHAI 718
 DB 667 VASRVYARRIGGCGMAFSGLGTQOALERIGRFAGALISVNGSPRVVAVAGSGDLDEL 726
 QY 719 AAMAAARGARTKALHVSFHSPLMAPLEAFGRVAESVYRRPSIVLVNLSGKACTDE 778
 DB 727 IAECEAEAKKARIIPVDVASHSPQVESLKEELLTELAGISPVASADVALSTTGGQPIDTA 786
 QY 779 VSSPGVWZHAREVVPADGVKALHAAGCTFYEVGPKSTLGLVPAKMPDARPA----- 833
 DB 787 TMDTAVMYNLEBOVAFODATROLACAGFAFEVFPVHVLTVGIGIATLDSALPADAGAC 846
 QY 834 LLAASRAGDEPRTVLEALGLMAVGLVSMGLFBSGGRVPLPTVPMORERWIDTKA 893
 DB 847 VGTGLRRDGGGLADFTALGEAFAOGVEVDWSPAF--ADAPVELLPVPPQ--RYMLPIPT 903
 QY 894 DDAAR--GDRAPGAGHDEVEGEAVRGDRRSARLDHP--PPE--SGRREVEAAGDR 946
 DB 904 GGRARDEDDMRQYVWREAMEWESASLAG--RVLLVTGPVSELSDAIRSGLESGAT 960
 QY 947 PERLEID-----EFGVLHL-----VLRVTERA--PGLGEVEI--AVDAAG----- 984
 DB 961 VLTCDVESSTIGTALAADVTALSTVGAAPPHGAVPSPDLALVALGAAGVEADL 1020
 QY 985 --LSFNDVOTALGMVDDLPKGNPPLLIGGECAGRIIVAG-----GVNGVLVAGPVIA 1037
 DB 1021 WULTRIAVAVADELVYD--PAQ-----AMVGG--LGVVGIEDQGRGGLVDLV----- 1065
 QY 1038 LSGAGFATHVTTSAALV--PRPOLSAIEAAMPVAYLTAWVALDRIARLQ--GERVLTH 1095

DB 1066 ---DADAASIRSLAAVLAADPRGEEQVAIRADGIKVARLVAPAPARATHLEPLAGVLT 1122
 QY 1096 AATGVGLAAVQWAGHVAEYHA-----TACTPERGATLESGLRVY-----VDS 1140
 DB 1123 GGTGIGIAGHLAARLSGAE--HLVLLGRGADAPASELEBELTALGCGVTTAACDVADR 1181
 QY 1141 RSDRFVADVAMTAGEVDVVLN---SLGSELJKSFNILRSRFPVELGKRDVYADNQ 1196
 DB 1182 ARLENVLAEEAAEBRTYSAVHHAAGVSTSTLDD-----LTAETFEI-----AD-- 1227
 QY 1197 LGLRPLRLNLSFLVDLRGMLEKPARVALLLEELGLIA---AGVTPPTATPLIAR 1252
 DB 1228 -----VKVGVTV-----NDELCPDIDAFVLFSSNAGVMSPELASYAAN 1268
 QY 1253 V-ADAFRSMAQ-----OHGKVLVLTGDP-----EQOIRI 1282
 DB 1269 AFLDGPAPARAREGAPVTSIAMGLMAGQNMADGEGEYLRSGGLRAMDRAVEELHTL 1328
 QY 1283 PTHAGAGSTGDRD-----LIDLASA-----APARAAA----- 1312
 DB 1329 -DHGQTSVVDMDRRFVELEFTAARHRLPFEIAGARAEAROSEGRLAQLAALCD 1387
 QY 1313 -----LEAFLETOVSQVL--RTPEIKVGAELFTRLGDSLMAVELRNRIEASLKLKST 1365
 DB 1388 GREHEHLALIRAEVAAVVGHDGDAIDRDRAFRDLGFSMTAVDLRRLAIVTGVEAA 1447
 QY 1366 TPLSTSPNIALAQLDLALATLSLERYAAE-----NLKAGVON 1405
 DB 1448 TVFDPHPTTLRDLHYLERLVGAABAEQAALVREVPKADDPFIAVGMACRFPQVYN 1507
 QY 1406 -----DFVSSG-----ADDMELIAL 1421
 DB 1508 PGLMEFTVGGSDATVEMPTGRGMDLAL 1536

RESULT 5
 MCAS MYCBO
 ID MCAS MYCBO STANDARD; PRT; 2110 AA.
 AC 002251;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Mycocerosic acid synthase.
 GN MAS.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxId=1765;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=92406887; PubMed=1527058;
 RA Matnur M., Koiatukudy P.E.;
 RT "Molecular cloning and sequencing of the gene for mycocerosic acid
 synthase, a novel fatty acid elongating multifunctional enzyme, from
 RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guérin.",
 RL J. Biol. Chem. 267:19388-19395(1992)
 CC - FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
 CC METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
 CC FORM MYCOCEROSYL LIPIDS.
 CC - COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETINE.
 CC - SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
 CC ARRANGEMENT.
 CC - SUBCELLULAR LOCATION: Membrane-associated.
 CC - SIMILARITY: Contains 1 acyl carrier domain.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M95808; AAA25369.1; -
DR PIR: B44110; B44110.
DR HSSP: P73283; 1E5M.
DR InterPro: IPR001227; Ac transferase.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; ketoacyl-synt.
DR InterPro: IPR006163; PP_bind.
DR InterPro: IPR006162; Ppanene_attach.
DR Pfam: PF00698; Acyl_transferase.
DR Pfam: PF00107; Adh_zinc_N.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR Pfam: PF00550; PP-binding; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
DR PROSITE: PS00066; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS50075; ACP_DOMAIN; 1.
KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
Transferrase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT DOMAIN 1 430 BETA-KETOACYL SYNTHASE.
FT DOMAIN 533 852 ACYL TRANSFERASE.
FT DOMAIN ? ? ENOYL REDUCTASE.
FT DOMAIN ? ? BETA-KETOACYL REDUCTASE.
FT DOMAIN 2026 2096 ACYL CARRIER (ACP).
FT ACT_SITE 177 177 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 623 623 ACYL TRANSFERASES (BY SIMILARITY).
FT NP_BIND 1561 1578 NADP (NR).
FT NP_BIND 1765 1780 NADP (NR).
FT BINDING 2059 2059 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 2110 AA; 225577 MW; C658215D7155300 CRC64;

Query Match 25.7%; Score 1852.5; DB 1; Length 2110;
Best Local Similarity 29.5%; Pred. No. 6; Se-90;
Matches 52; Conservative 176; Mismatches 519; Indels 555; Gaps 30;

QY 13 PIAIVGASCRPLPGGVIDLSGFWTLLSGSRDTYGRVPAERWDAAMFDPDPAEGKTPVTR 72
DB 7 PVAIVGCGCRPLPGGINSPPDKLMESILRGDDLVTEIPDRWDADDYDPEPGVPRGSVSRW 66
QY 73 ASFLSDVACPDASPFQISPREALRMDPAHLLLEVCHEALENNAIASALVGTETGFTG 132
DB 67 GGLFDVAGFDAPFQISEREATSIDPOQLLETSEALIEHAGLDASLAGSSTAFTG 126
QY 133 IGSEYEALPQATASAEIDAHGGL-----GTMPEVAGARISYALGRLGPCVAVDYAT 185
DB 127 LTHEDYIVLTFTA-----GGLASPYVVTGLNNSVVASGRITAHITGLGPRMPTFTAC 177
QY 186 SSSLIYAVHLACQSLRSGECSTALAGVSLMLSPSTLVMLSKTRALARDGCKAFSAEDG 245
DB 178 SSGIMAVHLACRSIHDEBDLALAGCAVLLLEPHACVAASAGMLSTGRCHSPDADAG 237
QY 246 FGRGEGCAVVLKRLSARADGRIILAVIRGSAINHGCASSGLTPNGSOEIVLKALA 305
DB 238 FVSEGGCAVVLKRLPALRDGNRIFAVVRGTATNQGRTETLTMPESDAQVAAYRPLA 297
QY 306 DACCAASSVGVENAGTGTGLGPRIEIQALNAVYGLGRDVATPLLISVKNLGHPEYAS 365
DB 298 AAGVQPTVGVENAGHTGTPIGDPIEYRLARVYAG-----TPALGSAISMNCHSTASA 353
QY 366 GITGLKLVLSLOHGQIPAHLLHAQALNPRISWGDRLITVTRATPMDMN--TPRAAGVS 423
DB 354 GTVGLIKAILSLRGGVVPRLHFNRLPDELSDVETGLFVPOAVTPMNGDHTPKRVAVS 413
QY 424 SFGMSGNNHVVLEEAATCTPPAPRRPAE-----LLVLSAKRTASALDAQAARLRLHET 479
DB 414 SFGMSGNNHVAIVEEAPAEASAPESPDAEVPRLFMSTSSDALARQOTARQATVVEE 473
QY 480 YPSQCL--GVAFSLATTRSAHEHRLVAATSR-----EGLEPAALDAAGQTSFGAVRSI 533
DB 474 H-QDCVAASDLATYTLARGRRHRRPRTAVVAAANLPELVEGLREVADGALIDAANG----- 527
QY 534 ADSSRGTLAFLFTGQAGQTLGMRGLYDVMSAFREAFDLCVRLFNQELDRPLREVMMAEP 593
DB 534 ADSSRGTLAFLFTGQAGQTLGMRGLYDVMSAFREAFDLCVRLFNQELDRPLREVMMAEP 593

DB 528 -HGDRGPV-WVPSGQGSQWAMGTOLL-----ASEPVFAATI-----AKLEPVIAESS 573
QY 594 A-SVDAALDDQTAFT-----OPALFTFEYALAA-LWRSWGVEPELVAGHSIGELVACVA 646
DB 574 GRSVTAITRAQCTVTYIDKVPFAVFAVQALAAITMETQTVGRPGAIVAGSMESAAVA 633
QY 647 GVFSLEDAVFLVAARGRLMOALPAGAVNSIEAREADV-AAAVAPHAASVSIAAVAPQ 705
DB 634 GALTSEDAARVICRRSKLMTRIAGAGMGSVLEPAQVNSELMARIGIDDVVVSVASPOS 693
QY 706 VVIAGAGQVYHIAAAMARGARTKLVHSHAFSHSLAMPLEAFRVAESVSYRPSIV 765
DB 694 TVVGSTSDYVRDLIARBEQRDVAREVAVDVASHSPQVDPILDDLAALADALAPMTKPV 753
QY 766 LVSNLSGKACTDEVSFGYVVRHAREVREPADCVKALHAAGAGTFEVEGPKSTLGLVPA 825
DB 754 YISATLFPDRREQVCAGAYVNDLNRVTQFPAAAVQAAAMDGRVFAVDVSPHLLTHAVQ 813
QY 826 C-MPDARPALLASRAGREDEPATVLEALGLMAVGLVSWAGLPPSGRRVPLPTYPQ 883
DB 814 TGRSLMSVAALAGMRREQPLPHGLGLLTLEHRAQALDYSALYPA-GRLVDAFLPAWT 872
QY 884 RERYTID----- 890
DB 873 HAFLLFIDDDGOEQORAGACTIVPHLGSVHLTEPERHVNQGVTSVLSWLSOHVH 932
QY 891 ----- 890
DB 933 NVAALFGAAYCEMAALAAAEVGEAAEVDTTFEQMLLDEQTPIDAVASIDAQGVNFT 992
QY 891 -----TKADDAARGDRAPG----- 905
DB 993 VETNRDGETTBHATATLRAAEDDCPPGYDITALLQHPHAVNGTAMRESFAERGVTIGA 1052
QY 906 -----AGH-- 908
DB 1053 AFGGLTAAHTPRRPRRCMPRRCRRSSRAPTESTRCWTLVSSRARASRYGHW 1112
QY 909 -----DEVEEGC---AVRG--- 919
DB 1113 PAVAVGCAQACLRAYPQCPRLHAVDGLQRDPWEADLVLDHGVYLLAARGLRNG 1172
QY 920 -----GDR-----SARLDHP----- 930
DB 1173 TGTSEDERDRLVSERLTLTGMOQRALPEVGDEAGSWLIDPSNAVDPDMLASTLTDA 1232
QY 931 ----- 930
DB 1233 LKSHGQGTGECASLWSVODTPPNQAGLEKLSQLRGDGVVYGVPRVGDPEHSLIA 1292
QY 931 -----PESGRREKVEAAGDR----- 946
DB 1293 GREQVHLVRIITRELAEPFEGELPRLFVITRQAOIVVPHDSGERANLEQAGRLARVIS 1352
QY 947 ----- 946
DB 1353 EHPMLRTLLIDVDEHTDVERVAQQLLSGSEDEETARNGDMVYARLTSPPLGHEERTAV 1412
QY 947 -----PFLIEIDEPVLDHLVLRTERARAPGLGEVIEIVADAAGLSFNDVOLALGNVP-- 998
DB 1413 LBDHDGMQVQVRPPDLOTLFEVASDRVPPGGQIEVAVSMSSINFPADVLLAFGRFPIT 1472
QY 999 DDLPGKPNPPLLGGECAGRIVAVGCVNGLVVGGCVIALS--AGAFATHTVTSAAVLPR 1057
DB 1473 DD--REPO-----LGMPFVGVTYAVGEGVGHQVGDVGVGFSGEGCKRTPLTCANLAVTL 1526
QY 1058 POALSAIEAAMPVAYLTATYALDRITARLOPGERVLIHAATGCVGLAAVQMAQVGAEVH 1117
DB 1527 PGLTDEQAITATATATATWAGLNDLAQIKAGDKVLIHSATGCVGOAIIISARAKAEIR 1586
QY 1118 ATATGTEKRAYLESIGVRYSDRSRPFVADVAMTGGEGGVVNLISGELIDKSPNLL 1177
DB 1587 ATAGNPAKAMLRDMGVENHYDSRVSVEFAEQIRBDTDGVDIVNLSTGAAGRAGLELL 1646

QY 1178 RSHGRFVELGKRDVCYNOLGLRPFIRNLSFSLVDRGMWLEBPAPVRLLELLGLIA 1237
 DB 1647 PFCGRFEVEIGKADVGNTRGLGPEFRGRGLTFYLLDLALMSVTPDRVRELLATVFLTLMD 1706
 QY 1238 GVFTTTPPIATLPIARADAPRSMQAQOHGKVL 1271
 DB 1707 GVLTAAOCTHYPLAADAIRAMSNABEHGKVL 1740

RESULT 6
 PSB_MYCTU STANDARD; PRT; 1538 AA.
 ID PSB_MYCTU Q10978; OS3234;
 AC Q10978; OS3234;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phenolchicocol synthesis polypeptide synthase ppsb.
 GN PPSB OR RV2932 OR MTJ3002 OR MTCY338.21 OR MT0011.01.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinomycetidae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_Taxid=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekie F.,
 RA Badcock K., Bauman D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Suleon J.E., Taylor K., Whitehead S., Barrell B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,
 RA Bishop W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POLYPEPTIDE SYNTHASE INVOLVED IN PHENOLPTHIICEROL
 CC SYNTHESIS.
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE (By
 CC similarity).
 CC -1- Similarity: Contains 1 acyl carrier domain.
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL021070; CA15929.1; -;
 DR EMBL; AE007122; AA47329.1; -;
 DR PIR; E70874; E70874.
 DR TIGR; MT3002; -;
 DR TIGR; MT3002; -;
 DR InterPro; IPR001227; Ac transferase.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR006153; Pp_bind.
 DR InterPro; IPR006162; Ppantn_attach.
 DR Pfam; PF00698; Acyl_transf; 1.

DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF02801; ketoacyl-synt C; 1.
 DR Pfam; PF00550; pp-binding; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; FALSE_NEG.
 DR PROSITE; PS00075; ACP_DOMAIN; 1.
 DR Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
 KW transferase; NADP; Phosphopantetheine; Complete proteome.
 FT DOMAIN 1458 1495
 FT BINDING 1458 1458
 FT ACYL_CARRIER (ACP).
 FT PHOSPHOPANTHETHEINE (POTENTIAL).
 SQ SEQUENCE 1538 AA; 162527 MW; B5E2A2042AD00CC CRC64;

Query Match 22.1%; Score 1592.5; DB 1; Length 1538;
 Best Local Similarity 31.2%; Pred. No. 2.2e-76;
 Matches 488; Conservative 207; Mismatches 600; Indels 271; Gaps 44;

QY 3 DPEPIBAEDPFAITGASGRPLPGSYIDLSGFYTLLEGSMDTGRPAEMNDAAAFDDPP 62
 DB 25 DVSRIIAVAEPVAVVIGCRFPDGDGPFSEFDFVAGNNAISTVPADMDAEAFYHDP 84
 QY 63 DAGKTPVTBASFSDVACFDASPFGISPREALRMDPAHRLLEVCWEALENAAIAPASL 122
 DB 85 LTFGKMTTGWGFVDVAGFDABFFGCTPREAANDPOGRMLLEVAEMLEHAGLPPDGL 144
 QY 123 VGTETGVFIGPSEYEALPOATASAEIDAAGLGTWPSVAGRIAYALGLRCPVAVD 182
 DB 145 GGTIRIAVMGVFENYOSWL--AASPQNVDAVSGTGNASITVGRISYLLGLRPAVAVD 202
 QY 183 TAYSSIVAVHLACGSLASGEESTLACGSLMSLSPSTLVMKSKRALARDRCFAFS 242
 DB 203 TACSSSLVAVHLACGSLRRETDLAGVSTLLEPEQOIAISAGLSPQRCFAFDA 262
 QY 243 AGFPRGECVAVVVKRLSGARADDRILAIRGSAIHNHDSGSLTVNGSSQETVLKR 302
 DB 263 ADFVRGEGAGVVLKRLTDVARDQVLAIVRGSVAVODGNSNGVTAPNTAQCQVIND 322
 QY 303 ALADAGCAASVGVYEAAGTGTLLDPIEQALNAVGVIGRDVATPLLIGSVKYNLGRPE 362
 DB 323 ALRSGDVAPDSVNYVEAHGTGLDPIEFELALATYGHGDA--LGAVKTYIGHLE 379
 QY 363 YASGITGLKLVLSIQHGOIPAHMLAQLNPLISGDLRLTYTRARTPPDNTRRACV 422
 DB 380 AAAGIAGFKATLAQORATIPNLFHSQMNPAIDASTRFPTQNSPPTREGRRRAV 439
 QY 423 SFGSGNTAVHLEE--PAACTCPAPERPELLVLSARTSALDAOARLRDLHETYP 481
 DB 440 SSGFGGTAHVITIQGSELAPVSEGEEDTGVSTLVGKTAQRMAATQVLADWMEGPG 499
 QY 482 SC-CIGDVAFSLATTSAMEHRLAVALAATSRREGRLAALDAAGQTSPGAIVRSIADSRCK 540
 DB 500 AEVAVADVAVHTNNHRAQATFTGVVARDRAQALIGLALAAAGQAPGV--SHQGSPPR 558
 QY 541 -LAPFTGQAGTLCMGSLYDVMSAPREBAPDLCYRLFNQELDRPLREYMAEPASVDA 599
 DB 559 GTVFYSGSGSQWAMGRQLADDEBPAFAAVALPEPVFEQAGFSLRDVI--ATGKELY 615
 QY 600 LLDQAFQOPPLFTFEVYLAALMRSWGVEPELVASHSIGELVAAVAGVAFSLEDVFLYA 659
 DB 616 GIEQ--TQLGLIGQLTTLTLMRSYGVOPDLVISHSGEVAVAAGALTLAEGRVYA 672
 QY 660 AGRILMQLPAGANAVSIEAPEADVAAVAPHAASVSIAAVAPQVNVAGAGOPVHA 719
 DB 673 TRARLMAPLSGGGGVALGLDAAALEALIDY--PQVTGIVNSPQVTIAGTBOIDEI 731
 QY 720 AAMAAARGATKALHVSHPHSPPLAAMPLEAFGRVAVESYRRPSIVLVNSLGAQCTDEV 779
 DB 732 AEVRAQNFASRVNIEVAPHPNDALQPAKSELADLTPRTGTIISTVYADLHTOP 791
 QY 780 SSGVWVHAREVAFADGVVLAHAAG-----TFVEVGKSTLLGLVPCMPDA--RP 832
 DB 792 PDAEHWATNKNRPVFO--QAIASAGGADGAYTFTEISAPHLITQAIADTLEDAHRP 848
 QY 833 A-----LLASPRAGDEPATVLEALGLMANVGLVSWAGLPPSGGRVP----- 876

Db	849	TKSAKYISICLQDADDV-----TFRNLVTTADIANHPHCPEEPHP	895
Qy	877	LPTVYQGERYWIIDTKADDAARGDRRACGAGHDEVEEGAVGGRSARLIDHP--PPE	933
Db	896	IPFTWQHHTHHVIAT-----THSSTAPE	919
Qy	934	SGRREKVEAAGR-----PFR--LEIDEPGV-----LDHLVLRV	965
Db	920	DPGSKNVVNNQGSTISESRALDEWCHQLAMPIRPAVSADPSTAAMLVVDNELCHELARA	979
Qy	966	TERRAPGCGEVEIANDAGLSENDVOLAGWPDLPCKPMPRLTGGECAGRIYAAVGE	1025
Db	960	ADSRDSDISPPRLAAGSDPALLD--ALRGVDNVLYAPRPVGGELLDIESAYQVPHATR	1036
Qy	1026	VNGLVVGQPVIALSAGAFATHTTSA-----ALVLPRPQALSAT	1064
Db	1037	LAAMVAVSASATISPPKFLI-MTRNAQOPISEGDRAHPGHAIVLMGLGRSLALEHPEIMG	1095
Qy	1065	-----EAAAMVAVLYLTATYALD-----RIARLO-----PGERVLIHA--	1096
Db	1096	IDLDSDMPAEALVRRHV--LTTAAGTGDGEDOVYRSGARHVRLORRTLPGRKPTVLTNDAS	1153
Qy	1097	-----ATGCVGLAAVQMAQHVGAE--VHATAGTP-----EKRAYLESLGYRVYS-----	1138
Db	1154	QLVITATIGNIGPHILIRQLARWGAKTIVAMARKPGALDELTCGLAATGTDLIAVAADATDP	1213
Qy	1139	---DSRSDRFVADVRAWTG-----GEGVDVYVNSLSGELIDKSPN-----LLRSHGRVE	1165
Db	1214	AAMQTLPRFGFELPRLEGIVLAAPAGRPALLSEMTDDVTTMPFRKDALALLHRRSLK	1273
Qy	1186	-----LGKR-----DCYADNQLGRPFELR-----NLSTSLVDLR	1214
Db	1274	SEVRHFVLFSSVSGLLGSRMLAHTYATSAFLDSSFAGARTGTLPATVDMGLMXSLADVQ	1333
Qy	1215	--GMMLEBPAPARVALLLEEL-LGLIAGVFTPPITLPIAR---VADAPRSMQAQHLG	1267
Db	1334	KDATOISABSGIQMADEVAIGALYLVMKPPDAAVATVVAADWPLAARVTRGALRIVD	1393
Qy	1268	KLVLT---LGPBEVOIRIPTHAGAGPSTGDRDLT-DRLASAPAPARAALAEAPLRTQVSQ	1333
Db	1394	DLLPAPEDVGKSESEFR--TSLRSCPAERBRMULMDHVGALAAATVWG-----	1438
Qy	1324	VLRPEIKVGAELATRLGQMSLMAVELRNRIEASLKLKLTSTFLSTSPNIALLAQNLID	1393
Db	1439	--MPTEPLDPESAGFEQGLQMSLMSVTLQRLALSLSLGEFLPASVVEDYPTV---YSLTD	1492
Qy	1384	ALATAL 1389	
Db	1493	YLATVL 1498	

RESULT 7

PPSA_MYCTU

ID_PPSA_MYCTU STRAND: PRT; 1876 AA.

AC Q10977; 01-OCT-1996 (Rel. 34, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Phenolphthalein synthesis polyketide synthase ppsa.

GN PPSA OR RV2931 OR MT3000 OR MTY338.20.

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.

NCBI_TaxID=1773;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H3TRV;

RA MEDLINE=98295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies K., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroyd S.,


```

Qy 118 ABSALVGETGVFIGIGPESEYALPQATAS--AEIDAHGLGTMSVGAGRISYALGJRG 176
Db 208 PGGLTFRSATGVFAAGACISEYGA---MASADLSQYDGMNSGGAMSIIANRLSYFLDLRG 264
Qy 177 PCVAVDTAVSSSLVAHVLACQSLRSCECTALAGVSLMSPSTLVMLSKTRALADGRC 236
Db 265 PSVAVDTACSSSLVHILACQSLRTQDCHLAIAAGNLLSPAVFPGFQVQLASPTQC 324
Qy 237 KAFSAEADFGEGGCAVVLKRLSGARADGRLLAVIRGSAIINDGASGGLTVPNGSQ 296
Db 325 RAFDITADGFVBEAGGVVVLKRLIDQRDGRVLAIVGSAVNOGSRNGLMAPRPAQ 384
Qy 297 ELVLKRALADAGCAASVGVYEAHGTGTGADPIETIOALNAYGIGRDVATPLLIGSVY 356
Db 365 MAVLBAAYTNAGMQSEVDYEAHGTGLIGDPIERBALGTVLGRGPRDPSLLIGSVY 444
Qy 357 NLGHEPYASGITGLKLVSLDHGQIPAHILHQAOLNPRISMGDLRLTVTRATTPMDMT 416
Db 445 NLGHEEAAAGIAGFTKTVLAVOHGQIPRNOHFETANPHIPTDLMKVVDTQTEWPAH 504
Qy 417 PRAGVSPGMSGTAAHVLEEARATCTPRAPERPA-ELVLSARTASALDAQARLD 475
Db 505 PRAGVSPGMSGTAAHVLEEARATCTPRAPERPA-ELVLSARTASALDAQARLD 564
Qy 476 HLETPSQ-CLGDVAFSLATTRSAMEHRLAVATSRREGIRALDAAAGQTSFGAVRSIA 534
Db 565 WMEGADVALADVAHTLNHRSRQKFTVVARBRTOAIAGLRBALAAGHAPGVVNP-A 623
Qy 535 DSSRKG-LAFLFTGGAQTLGMRGLYDVMSAPREAFDLCVRLFNOELDRPLREYMAEP 593
Db 624 DSSPFGTVFVYSGRGSQVWAGMGRQLADEPFAAAVALEPEVFEQAGFSLHDVL--- 679
Qy 594 ASVDAALDQTAFTQPALTFEYALALMRSGVBEVLVAGHSIGELVACAGVPSLED 653
Db 680 --ANSEELVGIEQIQGLIQMLATLWCISYGVREDLVIGHSWEGVAAVAAGALTPEAE 737
Qy 654 AVFLVAAGRLMQUALPAGG--AMVSIPEARADVAAVAPHAASVIAANAPDOVVIAGA 711
Db 738 GLRVATSRMLAPLSSGGGMALLEADATTEALADFP--QVTLGINSRQVVIAGP 794
Qy 712 GQPVHIAAAMAARGARTALHVSNAFHSPLMAPMLAEARVAESVYSRPSIVLSNLS 771
Db 795 TEOIDELIARVRAQNRFAFRVNI EVAHPNAMDALQPARSELDLTPRTPIGISTVY 854
Qy 772 GKACTDEVSPPGYWRHAREVVRFADGVKALHAAGG-----TPREVCP-----KSTLLG 821
Db 855 ADLHTQPVFDAEHMWTNMENPVRFQ--QALASAGSGADGAVHTFELISAHPLTQALID 911
Qy 822 LVPAICMPDARPALILASSRAGDEPATVLEALGMLAVGGLVSMAGLFP-----SGGRVP 876
Db 912 TLHSAQPGARVYSLGTLQRTDQDVTFRTNLAKAHTI-----HPHTHPREPHRP 962
Qy 877 LFTYVQWRERYWIDTKVADDAARGDRRAPAG-----HDEVEEGAVRGCDRRSARL--DH 929
Db 963 IPTTWMQHTRHMITTKYRPAAGVGS--APRAGTLLGQHTTVATVSPSHLMOALAPDA 1020
Qy 930 PPEBSGR-----EKTEAA-----GDRPFLELDERV 957
Db 1021 KPYQGHFHOVEVVPASVVLHTLILSAATELGYSALSEVRFOPIFADRPRLIQV---V 1076
Qy 958 LDHLVLVTERA-----P 971
Db 1077 ADNRAISLASSPAACTPSDRWTRHVTAOUSSPSSASLBNHHRANQOPPRAROLDLP 1136
Qy 972 GIGVEIADAAGLSF-----NDVOLALGVPPDLDFKRNPPILLGEGCAGRTV 1020
Db 1137 DLAEILARKGIDGLPEFSWTVASMTQHSNLTVAIDL-PEALPEGSTGPLL--DAAVILA 1192
Qy 1021 AVGEGVNG-LVYGGPVIALSAGAFHTVTSAL-----VLPRQALSIEAA 1068
Db 1193 ALSDVADSRLLVPASIEQISLGDVVTVGRSSVTLNRTAHDDGITVDVTVAHAGEVPSLS 1252

```

```

Qy 1069 M-PVAVLTAWYALDRIRALQGERVYLHAATGVG-----LAAYVM-----AQHV 1112
Db 1253 MRSILRYALDPLDGLD-VGSAQP-----PASTGPYAVCDATNFVHTIMQGTVDATHP 1305
Qy 1113 GAE--VH-----ATAGTEPKRAYLESIGVRYVSRSRSPFVAD----- 1148
Db 1306 GAEQVTHGPAVLIIGDDGALCETLEGAGYOPAVMSDQVARSRYVYVADSDPAGADETD 1365
Qy 1149 ---VRAVTGEG----- 1157
Db 1366 VDFAVRTCTETIGLVRTLAEADADKPAALMILTRGVHVSVPALROSFLMGAGVIAAE 1425
Qy 1158 ---VDVUNLSLSELIDKSEFLNRSHGRFVELKRD----- 1190
Db 1426 HPELWGLVLDALINDDGEFGPALLAEILAKPSKI-LVRDQGVLAAPALAPYRGPARKS 1484
Qy 1191 ---CYVD-----NQLGL-----RPFRLNLSFSLVD----- 1212
Db 1485 LQCRPDAAVLYTGGIGALGLMADWLDRGAHRLVLTGRTPLPERRDWQDLDTLRLRR 1544
Qy 1213 ---LRGMLEPRAR-----VRAL-----EELGLI----- 1235
Db 1545 IDALRALEMKGVTVEAVADVGCREDVQALLAARDGAAPIRGITTHAAGITINDOLVTSM 1604
Qy 1236 ---AAGVFTPP-----IATLP 1249
Db 1605 TGDVAVQVWMPKIGGSQVLDHAPPGSVDFYFLTASAGIGFIPQGSYAAANSYLDALA 1664
Qy 1250 IARVADARSM-----AQOHLGKLYLTLDPE----- 1277
Db 1665 RARQOGCHTMSLDVAVMRGLAADAQVSEELARMSRITPSEAFTEAFVVDGYDVA 1724
Qy 1278 --VOIRIPTHAGPSTGRDLDRLASAPAR--AALBAFLRTOVSOVLRPEIKVG 1333
Db 1725 QAVVMPMPAGADSGGNAYVLLPRKMSVMAATEVNSELOGLRRTIAELIRVEKEID 1784
Qy 1334 AEALFTRLGMSLMAVELRNRIEASLKLKSTFTEPSTSPNIALAONLLDALATLSER 1393
Db 1785 TDRPFAELGLNSLWMAIRREAEQVGELEATVLFNHPYKSLASVY-----AKR 1835
Qy 1394 VAENLRAGVQNDVYSSGA 1412
Db 1836 VAPHDVQODNQISALSSSA 1854

RESULT 8
WA EMENI
ID WA EMENI STANDARD; PRT; 1986 AA.
AC Q03149;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Confidial green pigment synthase (EC 2.3.1.-).
GN WA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TextID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93101122; PubMed=1465094;
RA Mayorga M.E.; Timberlake W.E.;
RT "The developmentally regulated Aspergillus nidulans wa gene encodes a
RL polypeptide homologous to polyketide and fatty acid synthases."
RL Mol. Gen. Genet. 235:205-212(1992).
CC -1- FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
CC INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
CC BY CONIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
CC ASSEXUAL SPORES (CONIDIA).
CC -1- COTRANSLATION: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES
CC (POTENTIAL).
CC -1- PATHWAY: Conidial green pigment biosynthesis.
CC -1- SIMILARITY: Contains 2 acyl carrier domains.

```


RESULT 9
MSAS: PENPA STANDARD; PRT; 1774 AA.
AC P22367;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 6-methylsalicylic acid synthase (EC 2.3.1.165) (6-MSAS).
OS Penicillium patulum (Penicillium griseofulvum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; microsporitic Trichocomaceae; Penicillium.
OX NCBI_TaxID=5078;
RN (1)
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=DSM 62862;
RX MEDLINE=91006137; PubMed=209605;
RA Beck J., Ripka S., Siegener A., Schlitz E., Schweizer E.;
RT "The multifunctional 6-methylsalicylic acid synthase gene of
RT penicillium patulum. Its gene structure relative to that of other
RT polyketide synthases.";
RL Eur. J. Biochem. 192:487-498 (1990).
CC -1- FUNCTION: This multifunctional enzyme is a polyketide synthase.
CC It catalyzes a total of 11 steps by seven different component
CC enzymes, in the biosynthesis of the antibiotic patulin.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + 3 malonyl-CoA + NADPH = 6-
CC methylsalicylate + 4 CoA + 3 CO(2) + NADP(+).
CC -1- PATHWAY: Patulin biosynthesis.
CC -1- SUBUNIT: HOMOMULTIMER.
CC -1- INDUCTION: IN THE LATE LOGARITHMIC GROWTH PHASE.
CC -1- SIMILARITY: WITH MOST POLYKETIDE SYNTHASES, SPECIALLY WITH RAT
CC FATTY ACID SYNTHASE, AND WITH OTHER ENZYMES SUCH AS LIPASES AND
CC THIOLEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55776; CAA39295.1; -.
DR PIR: S13178; S13178.
DR InterPro: IPR001227; AC transferase.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR006163; Pp-bind.
DR InterPro: IPR006162; Ppantenn-attach.
DR Pfam: PF00698; Acyl_transf_1.
DR Pfam: PF02801; ketoacyl-synt_1.
DR Pfam: PF00550; pp-binding_1.
DR PROSITE: PS00012; PHOSPHORANTHENEINE, 1.
DR PROSITE: PS00606; B KETOACYL SYNTHASE; 1.
DR PROSITE: PS00075; ACP DOMAIN; 1.
KW Multifunctional enzyme; Transferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine.
FT DOMAIN 186 238 ACYLTRANSFERASE (AT).
FT DOMAIN 642 676 ACETYL/MALONYL TRANSFERASES.
FT DOMAIN 1403 1450 2-OXOACYL REDUCTASE.
FT DOMAIN 1700 1769 ACYL CARRIER (ACP).
FT NP_BIND 1424 1424 NADP (POTENTIAL).
FT ACT_SITE 204 204 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 653 653 MALONYLTRANSFERASE (BY SIMILARITY).
FT BINDING 1732 1732 PHOSPHOPANTHENEINE (BY SIMILARITY).
SQ SEQUENCE 1774 AA; 190732 MW; 05ED5DD10863F938 CRC64;
Query Match 16.3%; Score 1177; DB 1; Length 1774;
Best Local Similarity: 25.6%; Pred. No. 2e-54;
Matches 458; Conservative 210; Mismatches 655; Indels 466; Gaps 56;
QY 14 IIVAGSCLPGGVLDISGFWTLBGSRTVGRVPERWDAAMPDDDAAGKTPVTRA 73
DB 35 VAVVAGACRVAGNHNHPILLQSLQSKAMGEIIPMRWEPPYRRDARNEKELKOTTSRG 94

QY 74 SELSDVACFDASFFGISPREALRMDPAHRLLEVCMEALENAAIPASALVGTETGVFIGI 133
DB 95 YFLDLREDFDQCFISPKBAEQMDPQQRVSLEVASLEADGIAKSLSGSDTAVFCV 154
QY 134 GPSEYEA---LPQATSAETIDAHGIGTTPSVAGAGISTALGRGCVAVDTYSSSL 189
DB 155 NSDDYSKVLDELFP-----VWEAMMGIGTAYCGVPNRISYHLNLMGSRSTADDAACASLL 208
QY VAVHLACOSLRSGCESTLAGVSLMLSPSTLVLWLSKTRALARDCRAFASEADQFGRG 249
DB 209 VAIHGVQAIRLGESEKVAIVGVNALCGPGLTRVLDKAGALISSDSCSCSPDDAHGYARG 268
QY 250 EGCVAVVVLRSLSGARADDRILAVIRGSAINHDSAGSLTPVNGSSOEIVLKRALADAGC 309
DB 269 BGAGALVLRSLRALLDHDNLAVIKGSAVODGKTGNIMAPNSVAQQLAANNALSAANI 328
QY 310 AASSVGYEANGTGTLLDPPLEIQALNVAVYGLGRDVATPPLIGSVKTNLGHPEVASGITG 369
DB 329 DPHTVRYEAAHATSTPLDPTETISAIASVYCADRPADPCYIGSIKPNIGHLEAGAGVWG 388
QY 370 LKXVLSLQHQIIPAHILAQNLNPRISMGDLRLTVTRARTPMDNTPRRAGVSSFGMSG 429
DB 389 FLKAVLAIQKGLPPOANLTKINSRIDKTAGVKVQATPWPESDPIRRAGVCSYGYG 448
QY 430 TNAHVLEE-APATCTP---PAPERPAELVLISARTASALDAQAPARLDHLETY-PSQC 484
DB 449 TVSHAVIEFEPPILOPDLNGAVSGFG-LILSSPOEKRLALQKTRIDMTATGAKOHN 507
QY 485 LGDVAFSLATTSRAMEHRLAVALATSRERGLRALDAAQG--QTSFGAVRSIADSSRGKLA 542
DB 508 LSDIITLTLATRRDHDHVRALVDDYRDABEQVLSLAVGVHTFTQSRVLSGDISKDV 567
QY 543 FLFTQGGQITLGMGGLVDWVSAPREAFDLCVRLFNQELDPLREVMMAEPASVDAALD 602
DB 568 WVFSGHQWDMGKQL-----IHNPVFPAI-----QPLDLIAEGLSPTEILR 614
QY 603 QTAFA-----TOPALTFEYALAAALRSWGVPELVAGHSIGELVACVAGVSELDVAFL 657
DB 615 TQDFSSSRVQILTYVMQIGLSALQSGITPPQAVIGHSVEGIELASVAVAGALSPBEGALI 674
QY 658 VAARGLMQALPAGAMVSIAPADVAAPAAVAPHAASVIAAVNAPDOVINAGAOQPVHA 717
DB 675 VTRRALTLRYQWKGGMILVLPSEATEEILIGSRSDLV-VAIDSSPSCCVVAGDELVAE 733
QY 718 IAAAMAAGARTKALHVAHSHAPLMALPFAFGV-AESVYRPSIVLVSNLSGKACT 776
DB 734 TAEALKARGVKTFTYKSDIAFHSPPTLNGLVPLRDVLAEITLSPVSPNVKLYSTALADPRG 793
QY 777 DEVSSPGYVVRHAREVVFADGVKALHAAAGATFEVGPCKSTLLGLVPACMPDAR----- 831
DB 794 QDLRVEYVWAGNMVNRVLTSAVKAAYVEDGRLFEVSTHPVSHSINETLMDAGMEDPA 853
QY 832 --PALLASSRAGDEPAT--VLEALGGLMAGVGLVWAGLFPSSGRRVPLPTYPQOREY 887
DB 854 VIPTLL-----KKPTEKHILHSIAQHCAGAEVWMAQNP-GRWATGVPTTWMHKPI 906
QY 888 WIDTADDAAGDRAP--GAGHDEVEEG-----AVRGDR--RSARDH--PEPE 933
DB 907 W-----RKIETAPLHTGLTHD-VEKHTLLGORIPVPGDITVYVTRLDNDTKPFPG 956
QY 934 S---GRREKVEAG-----DRPFREI----- 952
DB 957 SHPLHGTIVPAAGLINFLLKGTGQMQLQNVVLRVPVAINAPRSQVVVVODQVAVSRL 1016
QY 953 --DEPGVLDHLVRLT-----ERRAPGLG-----VEJAV 980
DB 1017 IPSEBSQDDDDASWYTHTTAYWDRKVASSEDRIDPAVKSRLVTXLADNFSDYDKGV 1076
QY 981 DAAGLSF-----NDVOLALGM-----VPDDL- -GK- -RNP 1007
DB 1077 SAMGPFMAVHTHYRDKEMLARLVDPNPALISGADLPMDSSWAPVLAATAVSSTIFPTP 1136

```

QY 1008 PLL-----GGEAGRIIVAVG----- 1023
DB 1137 ALAMPQIERVEFTSQDPKISMLVQGEASDVPSHSVSEAGEVLAFTKMPRESI 1196
QY 1024 EGVNGL-----VVGQPVIALSAGAFATHTTSAALVPPQALS-AIEAAMP----- 1070
DB 1197 EGPVGSVGSMEIIVHQ--IAMPATPABEPISIEITVLVSPDATTALVAASIPTRVNSF 1254
QY 1071 -----VAYLT-----AMVMDJIALQPE-- 1090
DB 1255 QFSSTQEFFSNASSPLEKGTVTYITPGEVASLAEPVASESFTWMLLEIKETVNSGLP 1314
QY 1091 -RVLIHAATGCVGLAAVQMAQ-----HYGAEVHATAGT-----PEKRAYLESL 1132
DB 1315 IKFTLTANIGEOPTPALAQSPYGLARIASHBPLGLIDVEEVIPLSTMRYIQA 1374
QY 1133 GVRVSD--SRSDRFVADV-----AMTGEGCV----- 1158
DB 1375 DIRINDGIARTSRFRSLPRNKLPLASEGPRLLPREGTVYLTGGGLVGLLEVADFLVEK 1434
QY 1159 -----DVVNSLSGELIDKSFNILRSHGRFV-----ELGKRDQYAD 1194
DB 1435 GARRLLISRRALPPRRTWQVSEDIQPTI--AKIRLESRGASVHVLPLDITKPPAVEQ 1492
QY 1195 NQGLRPFRLNLSF-----SLVDLRG-----NMLERPARVRLLEELLGLIAAGVFTPP 1243
DB 1493 ----LTTALDRLSIPVQGVHAAGVLDNELVMQTTDAFNRYLAPKAGALALHEVFP 1548
QY 1244 P-----IATLPI--ARVADAF----- 1257
DB 1549 KSVDFEVMFSSCGNLVGTQASGSGNAFLDPLATRRALRGDAVASPQWTSWRGLDMGA 1608
QY 1258 -----RSMQAQHLGKL-----VLTGDPRE 1277
DB 1609 STDFINAELESKGTITVTRDEAFPAWQHLLAKYMDHGVILRSRAFEDEGPIPVSLINDIA 1668
QY 1278 VQIRIPTHAGSPSTGDRDLIDRLASAPARAALAEFLRTQVSOVLRTPEIKVGAEL 1337
DB 1669 VR-RVGVSTVNTSPRAAGSS--DAVPTSGPELK-AYDEKIRGCYAKVLQMTAEVNDVSKA 1724
QY 1338 FTRIGDMSLMAVELRNRIEASLKIKLSTFTPLSTSPNIALIALQNLIDALA 1386
DB 1725 LADLGVSVMVTTLRQLQTLKIAVPPTLWHSPTSLHVAWFAEKLA 1773

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: LA2766; AAC41675.1; -
DR EMBL: LA2766; AAC41674.1; -
DR PIR: T17490; T17490.
DR InterPro: IPR001227; Ac_transferase.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR006163; PP_bind.
DR InterPro: IPR006162; Phantne_attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl_transferase.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt; C; 1.
DR Pfam: PF00550; pp-binding; 1.
DR PROSITE: PS00975; Thioesterase; 1.
DR PROSITE: PS00075; ACP_DOMAIN; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; FALSE_NEG.
DR TRANSFERASE; Acyltransferase; Phosphopantetheine;
DR Multifunctional enzyme.
FT DOMAIN 374 805 BETA-KETOACYL SYNTHASE.
FT DOMAIN ? 805 ACYL/MALONYL TRANSFERASES.
FT DOMAIN 1714 1785 ACTL_CARRIER (ACP).
FT ACT_SITE 543 543 THIOESTERASE.
FT ACT_SITE 993 993 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 993 993 ACYL/MALONYL TRANSFERASES (BY SIMILARITY).
FT BINDING 1746 1746 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT_SITE 1937 1937 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 2109 AA; 230715 MW; CB701372A16D8551 CRC64;

Query Match 16.3%; Score 1173.5; DB 1; Length 2109;
Best Local Similarity 27.2%; Pred. No. 3.8e-54;
Matches 409; Conservative 222; Mismatches 625; Indels 247; Gaps 54;

QY 9 AADPPIATVAGACRLPGVYIDISGEFTLLGSRDITVGRPAERMDAAAFDPDPDAPGKT 68
DB 369 AGCKXALATVMSGRFPESPPTES-FWDLLYKGLDVCKEYPRRRWDINTHVDPGKARNKG 427
QY 69 PVTBASFLDVACPDASFPGISPREALRMDPARLLEVCMEALENAALAPSALVGT-- 126
DB 428 ATKMGCLDPSGDFDPFGIISPKXAPQMDPQRMALMTEMEARAGLVPTTPTSTQND 487
QY 127 -TGVFTIGPSEYEALPQATASAEIDAH---GGLGTMPVAGRIYSALGRGCVAV 181
DB 488 RIGVFHGVTSNDK---METNTAQNDITVFTIGNGNGFIP---GRINCFEFGAGSYTV 539
QY 182 DTVASSLVAVHLACOSLSGECSTYLAGVSLMSPSTLWLSKTRALARDGCAFA 241
DB 540 DTAQSSSLAIHILACNSLWRGDCDITVAGTWNATVYPPDHTGTDGKFFLSRTGNCCKPYD 599
QY 242 EADGPFRRGSCAVVVKRLSGARADGRIIAVIRGSAIHNHDAAGSLGTVPNGSSQETVLK 301
DB 600 KADGVCRAEGVGTVPFKRLLEDALANDPILGLVLDKATNSAMSESMTRPHGAQIDNMT 659
QY 302 RALADGCAASVGVYEAHGTGTLTGLDPIEQALNVAVG---LGRDVATPLLGIVKTNL 358
DB 660 AALNTTGLHNDSYIEHMGITGVQVDVAVMESVSVFAPSEFARADQPLFVGSAAKAV 719
QY 359 GHPEVASGITGLKAVLSLQHGQIPHL--HAQALNPR1-SWGDRLVTVTRAKTWPMW 414
DB 720 GHSEGVSGVSLIKVLMQMHDPIPHCGIKPSKINRNPFDUGABNVHIAFEPKWPRT 779
QY 415 NTFRRAGVSSFGMSGTNAVHLEAPATCTPPAPRPALVLSARTASALDAQAARL 474
DB 780 HTRRVVLINFSAAAGNTALIVDAPERHMPTEKDRSSHIVALSAAVGSAMKTUNLERLH 839
QY 475 DHELYPSQCLGDVAVSLATTRSAMERHLAVAAATSRREGRLAALDAQAQ--GQTSQAVR 531

```

```

Db      840  QYLLNPTDIAQLSYTTTARMHYLVHVSVTGASVEERTKLENAIONGVSRP----- 895
OY      532  SIADSSRKLAFLEFGAGQOTLGMGRGLYDWASAREAFDLCVRL-----FNGELDR 583
Db      886  ----KSKKILPAFGGGSQYATMGKQYDAPSPREDEKFDRLAQSHGPFSPFHVCTS 951
OY      584  PLREVMMAEPASVDAALDQTAFTOPALFTEFEYALAAWMSGVPELWAGHSIGELVAA 643
Db      952  PKGDVEEMAPVV-----QLATICLQMALTNMTSGIRPDVTGHSIGEFALL 1000
OY      644  CVAGFISJEDAVFLVLAAGRLMOALPRAG--AMVSIAPENDVAAVAAPHAASVIAVN 701
Db      1001  YAAAGLSASDVVYLVGORAEILQERCORGTAMLVKATPEALSQWIDH--DCEVACIN 1058
OY      702  APDQVVIAGQOPVIAIAAAMARGARTKALVSHAFSPILMAPLEAFGRABESVSR 761
Db      1059  GPEDTVLSGTTKNNVAEVRAMTNDGICKTLKLPFAFSAQVQPILDDEALQCATPAK 1118
OY      762  PSIVLVSNLSGKACTDE--VSSPGVWRHAREVVRPADVKALHAAG----AGTFEVGPX 816
Db      1119  POLILSLPLRTEIHGEGVTPSYVAQHCRHTVDMAQLRSAREKGLIDDKTLVIELGPX 1178
OY      817  STLGLVPCMPDAPRAL--LASSRAGDEPATVLEALGGLMANGLVSWA---GLPFGS 871
Db      1179  PLISGMVMTVLGDKISTLPTLAPNKA---IWPSTOKILTSVYTGGMINDMKYHAPFAS 1235
OY      872  GRVPLPYPMQREHYWIDTKAD---DAARGRRAPAGHDEVEGAVRGDRRSALD 928
Db      1236  QKVVDLPFYGMDLKQYVLPYQGDWCLHRHQDCKAPRHE-----IKTADVQ----- 1283
OY      929  HPPESSGREKEVAAGDRPFLEIDE---PGVLDELVL--RVTERRAPGLGEVEIYVDAAG 984
Db      1284  -VPEPS-----TPHRPSKLDPEKEAPEIKTTTHRVETTKLGLAT--LVETD 1332
OY      985  LSFNDVQ-LALGWPDLPKGRNPPLILGECAGIYAVGE--GVNGLVGQPVIALSAGA 1042
Db      1333  ISRKDVNGLARGHLVDGI-----PLCTPSFADIAMQGVQYSQRLRAHP---GAGA 1382
OY      1043  FATHTVTSALV---LPR---PQAL-SAIEAAMPVYLTWVYLDLIR-LQGEERVL 1093
Db      1383  IDGLVDVDMVVDKALVPHKGFPOLLRTTLTWMPPKAAATRSKVKFATYFADGKLD 1442
OY      1094  IHAA-----TGGVGLAAV-----QWAOHVGAEVHATAGTPEKRAYLESIGRVYVDSRS- 1142
Db      1443  EHASTVAFSTDQOLKSLRSVSEYKTHI-RQLHGHAKGQRMNRKTKYKLMSSMARF 1501
OY      1143  ----DRFYADVRAWTGEGVDVINSLSGEL-----I 1170
Db      1502  NPDMVLLDYLVLTNEAENEASGVDPSLSSSEGTFAAHPAHVDALITOVAGFAMNANDV 1561
OY      1171  DK-----SFNLL-----RSHGRVELEKRCVCYADNOL-----GLRPLRL 1204
Db      1562  EKQYVNVHGMDSFOYQPLDNLNKSQYVTKMGOA---KENDLVHGDVVVLDEQVAFPR 1618
OY      1205  NLSTSLVLDLRGMLEPARPARVALLLELLGLIAAGVFT---PPPIATLPIARVADAFSM 1260
Db      1619  GLTLRSVPRGALRVVLQTTVKADQL-----GKTMPSPPPTTTPISGYKRAKNOV 1672
OY      1261  ----AAQHLGKVLTLGDEVOIRIPTAGAGSPSTGDRDLRLASAAPARAAA-- 1312
Db      1673  SSOAIPAEATH-----SHTPPO-----PKISPV-PET-----AGSAPAAKGVS 1711
OY      1313  ---LEAPLRTQVSOVLRTPEIKVGAEL-----FTRLGMSLMAVELANRLIASLKLUS 1364
Db      1712  NEKLDAVNRV-VSE-----ESGIALEELTDSNFDMDGIDSLSSVIGSRPREDGLDLG 1765
OY      1365  TTF 1367
Db      1766  PEF 1768

```

```

FAS RAT
ID - FAS RAT STANDARD; PRT; 2505 AA.
AC P12785; 009187; 009190; 064717;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.39;
EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
GN FASN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89240686; PubMed=2717611;
RA Amy C.M., Witkowski A., Niggert J., Williams B., Randhava Z.,
RA Smith S.;
RT "Molecular cloning and sequencing of cDNAs encoding the entire rat
RT fatty acid synthase."
RL Proc. Natl. Acad. Sci. U.S.A. 86:3114-3118(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Liver;
RC MEDLINE=93075999; PubMed=1339331;
RA Beck K.F., Schreglmann R., Stachopoulos I., Klein H., Hoch J.,
RA Schweizer M.;
RT "The fatty acid synthase (FAS) gene and its promoter in Rattus
RT norvegicus."
RL DNA Seq. 2:359-386(1992).
RN [3]
RP SEQUENCE OF 75-2505 FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Mammary gland;
RC MEDLINE=89128431; PubMed=2915923;
RA Schweizer M., Takabeayashi K., Beck K.F., Schreglmann R.;
RT "Rat mammary gland fatty acid synthase: localization of the
RT constituent domains and two functional polyadenylation/termination
RT signals in the cDNA."
RL Nucleic Acids Res. 17:567-586(1989).
RN [4]
RP SEQUENCE OF 2085-2505 FROM N.A.
RX TISSUE=Mammary gland;
RC MEDLINE=88087240; PubMed=2891707;
RA Niggert J., Witkowski A., Mikkelsen J., Smith S.;
RT "Molecular cloning and sequencing of a cDNA encoding the thioesterase
RT domain of the rat fatty acid synthetase."
RL J. Biol. Chem. 263:1146-1150(1988).
RN [5]
RP SEQUENCE OF 1921-2324 FROM N.A.
RX TISSUE=Mammary gland;
RC MEDLINE=87246646; PubMed=3109907;
RA Witkowski A., Niggert J., Mikkelsen J., Smith S.;
RT "Molecular cloning and sequencing of a cDNA encoding the acyl carrier
RT protein and its flanking domains in the mammalian fatty acid
RT synthetase."
RL Eur. J. Biochem. 165:601-606(1987).
CC - FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF
CC LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH.
CC THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN
CC ACYL CARRIER PROTEIN.
CC - CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+) .
CC - CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC acetyl-[acyl-carrier protein].
CC - CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC - CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC [acyl-carrier protein].
CC - CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC - CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.

```

[illegible]

Db	50	RSGLKDLXSKFDSAF	FCVHPKQANTMD	PQRLLELVSEYAEIVDGGI	INPASILRGNTGVW	109					
Qy	132	GIGPSEFEALPQA	TASAEIDAHGGL	GTMSVGAGRI	SYALGLRGPCVADTAYSSGLVA	191					
Db	110	GVSSSEASEALSR	PEP-	-LIGYSWGCORAMMANR	ISFFDFGSPIDALDTCCSSLLA	167					
Qy	192	VHLACOSLRSGE	GTALAGVSMLS	STPLVWLSKTLAL	ARDCKFAESADGFCGEG	251					
Db	168	LONAYQARSCE	CPALVGGINL	KNTSVQFMKLM	SPDGTCSRFDSDSGNYCAEA	227					
Qy	252	CAVVVLRKLSGAR	ADRRILAVIR	GSAINHDA	-SSGLTVPNSSGEIVLKRALLADGCA	310					
Db	228	VVAVLTLTKKSL	AR----	RYVATILMAGNT	MDGCKEQGVTFPSSGAQSQLIRSLYQPGVA	283					
Qy	311	ASSVGYEANGT	TTLDADPIEIOAL	NAVYGLRDVA----	TEPLLGSVKNLGHPEVAGS	366					
Db	284	PESLEYIEAHOT	GVKBDP----	DELN----	GITRSLCAFROSPLLIGSTKSNMGPEEASG	337					
Qy	367	ITGLKVLVLS	OHGOI PAHL	HAOLNPNRI	-SWGSLRLTVYRARTPMDMTPRR----	AGV 422					
Db	338	LAALT	KVLLSENVA	PNLHFNPNRI	PALDGRLOV-----	DRPLPRGGIVGI 390					
Qy	423	SSFGMSGTNAVU	-----	EEAPATCTPPAPER	E-AELVLSARTASALDAQARLDHL	477					
Db	391	NSFEGGANVAV	ILPMTQQA	PA-----	PAPHALPHLLASGRITWBAVGLLEQGRHS	445					
Qy	478	ETYSOCLGDVA	BSLATTR	SAMERLVAAT	SRBGLBAALDAAGQTS	SGAVRSIADSS 537					
Db	446	Q-----	DIAF-----	VSMLDIATPTA	AMPFGYTVLVGEVHQ--	EYQOVVASQ 489					
Qy	538	RGLTAPLE	TCGACATL	MGRLYDWSA	PEAPRPLCYRLFNOELDRPLREVMABPASVD	597					
Db	490	R-PLMF	ICSGMG	TWRGMSLMRL	-DSFRST-----	LNSDELRPLGVKVSDDLSTD 542					
Qy	598	AALIDOTA	FTPALPTE	EYALALMS	RSWSEPELVASHGISIGELYAACVAGVFSLEDAVFL	657					
Db	543	ENTDDDI	VHSVSLTA	IOALIDL	ITSMGLKPDGI	IGHISGEVAGVADGQLSGREAVLA 602					
Qy	658	VAAGRILMO-	-ALPAGAN	VSIEAR	PEADVAANA	PHAAVSIAA-VNAPDOVITAGAOV 715					
Db	603	AYMGQOC	KIDANLPA	-GSM	AVGLSWECKORCP	-----GVVPA	CHNSDVTITISGPOAV 658				
Qy	716	HAIAA	MAABAR	KALVSH	-AFHSYLM--	-AP-ML	EAFGRV-----AESVYARP 762				
Db	659	NEPWEOL	KEVPAKEV	RTGLAHS	YFMEGIA	FTLLQAL	KVIRREPRPSARKWLSTIP 718				
Qy	763	SIVLVSN	LKACIDE	VSSPGYWR	HAREVVR	FADGVKAL	HAAGAGTFVEVGPKST-- 819				
Db	719	EACQOSSIA	-----	RSSAEYNN	NVLVSPLY	POEAL--	MWPEHAVLEIA	PHALLQAV 770			
Qy	820	--LGLV	PACMDAR	PALIAS	RACRDE	PAVLEAL	GLMAVGLVSNAGL	FPSGGRVY- 876			
Db	771	LKRGV	KPSC-----	TIIP	LMKR	DKDNL	LEFLT	LGKVHLTGIDIN	ENNALFPPEVFPVR 825		
Qy	877	-----	LPYQ	RRRWMT	KADDA	ANG--	-----	DRRA- 903			
Db	826	GTPL	ISPRIK	WDHSQ	TWDI	PVAEDP	PNGSSSSA	TYNNIDASS	SSSHYLVDCIDGRVL 885		
Qy	904	-----	PGAGH	-----	-----	-----	-----	-----	D 909		
Db	886	FP	GCYILV	WYKTL	ARSLS	LSLEET	PVVEFVNT	HOATIL	PRGTIVLEVRLLASHAFE 945		
Qy	910	EVEEG	AVRG-----	DRSAR	LDP--	PPES-----	-----	-----	934		
Db	946	VSDG	NLIVSG	KVYQW	EDPD	SKL	PDHPEV	P	IPASESES	SVRLTGEVYKELRLRGDYGP 1005	
Qy	935	-----	-----	-----	-----	-----	-----	-----	-----	934	
Db	1006	FGV	YEATLE	GEOK	ILMKN	WVTFMD	TMLQIS	ILGESK	OSLQ	PLRVTAIYIDPAHLQ 1065	
Qy	935	-----	-----	-----	-----	-----	-----	-----	-----	934	
Db	1066	KVYML	EGDO	VADT	TSRCL	GVTV	SGGVY	ISRL	OTTT	ATSRQEOBOLV	PTEKEVFTPHVE 1125


```

QY 935 -----GREKVEAAGRP----- 947
Db 1126 PEGLSASATLQKELQCKGLAKALQTKATQOGLKMTVPQGLDPCNGLRLLAAACQLQ 1185
QY 948 ---FRLEIDEP----- 955
Db 1186 NGNLQLELGEVLARELLPEPDLISGLINSQALKACIDTALENSTLKMKVVELAGSG 1245
QY 956 -----GTL 958
Db 1246 HLYSHISALLNTPMLQLEYTATDRHPQALKQVQTKLQGHDAVAGQWDSGPAPTNLQAL 1305
QY 959 D-----HLVLR----- 964
Db 1306 DLVNCNCALATLGDPAALADNMVAALKDGGFLMHVTLKGHALGETLACLPSEVQGPSPF 1365
QY 965 -----VTERRA----- 970
Db 1366 LSQEWESLPSRRKALHLVGLKKSFYGTALFLCRSLSPQDKPIFLPVEDTSFGVDSLKSI 1425
QY 971 -----PGLGE 975
Db 1426 LATSSQPVWLTAMNCPISGVVGLVNCJLKEKPGHRIRCIILSNLSTSHVPKLDPGSSB 1485
QY 976 VEIAYDA----- 982
Db 1486 LQKYLESDLVNVVYRDGANGARHHPQLEDDKEBEQTAHAFVNVLTGDLASIRWSSPLK 1545
QY 983 -----AGLSFNDVOLALG-NVPDDLPGK-PNPPILLGEGCAGRIYAVG 1023
Db 1546 HMQPESSSQAQCLTVYVYASLNFEDIMLATGKLSPDALPGKMASRDCMLMERSGR-DKCG 1604
QY 1024 BEVNGLVGQPIYALASAGAFATHTVTSAALVLRPQALSAIEAAMPVAYLTAMVLDLI 1083
Db 1605 KRWGVLV-----PAEGLATSVLLSPDLMDVPSSWTLERASVPVYTTAYVSLVYR 1656
QY 1084 ARLQGERVLIHAATGVGLAAVQMAQHYGAEVHAGTPEKRAYLES-----LGVRYVSD 1139
Db 1657 GRIQGETYLHISGSGVGOAAISLISLGCRTVFTVSAEKRAYIQAFPPOLDIDTSPAN 1716
QY 1140 SFSDFRVADVAMTGGEGVDVVLNSISGELLDKSFNLLSHGRFVGLGRDYADNQLGL 1199
Db 1717 SBDTSPEQVHLHTGGKGVDLVLSLAERKLAQSVCLAQHGFLEIGKFDLSNNHPLDM 1776
QY 1200 REFPLRLTSVLDKGMLEERARVALLBELGLIAAGVTFPPPIATPIARVADAFS 1259
Db 1777 ALFLKNVTGILLDLALFEGANDSWREVAELKAGIRDVVVKLCTVFPKQVDEDAFRY 1836
QY 1260 MAOAHGLGLVLTLDPEVOIRIPTHAGAGPSTGDRDLRLASAPARAALAEALFRT 1319
Db 1837 MAQKGIQGVLVQVREPEEMLP--GAQPT-----LISAIKTFCEBHSYIT 1884
QY 1320 -----OVSO-VLRTP-----IKVGAEL-----FTRLGMSLMAVELNRIE 1356
Db 1885 GGLGFGELRLARMLVLRGAORLVLTSGSIRGCVQAKHVRWRQGIHVLTSTNSVSL 1944
QY 1357 ASLKLKLTSTTFLSTSFNIALAQNLLDAL 1385
Db 1945 GARALIAEATKLGPGVGFNLAMVLRDAM 1973

```

```

RESULT 12
STCA EMENT
ID STCA EMENT STANDARD; PRT; 2181 AA.
AC Q12397;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative sterigmatocystin biosynthesis polyketide synthase (PKS).
GN STCA OR PKYST.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

```

```

OC Eurotiaceae; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 26;
RX MEDLINE=96202293; PubMed=8643646;
RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA Keller N.P., Adams T.H., Leonard T.J.;
RT "Twenty-five conserved transcripts define a sterigmatocystin gene
cluster in Aspergillus nidulans";
RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=95370159; PubMed=7642507;
RA Yu J.-H., Leonard T.J.;
RT "Sterigmatocystin biosynthesis in Aspergillus nidulans requires a
novel type I polyketide synthase";
RL J. Bacteriol. 177:4792-4800 (1995).
CC -1- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYKETIDE NUCLEUS OF
STERIGMATOCYSTIN FROM HEXANOYL-COA AND SEVEN MALONATES.
CC -1- CORFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES
(POTENTIAL).
CC -1- PATHWAY: Sterigmatocystin biosynthesis; first step.
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@ib-sib.ch).
CC
DR EMBL: U34740; AAC9191.1; -
DR EMBL: L39121; AA81586.1; -
DR HSSP: P25715; 1MLA.
DR InterPro: IPR001227; Ac transferase.
DR InterPro: IPR000794; Ketoacyl-synth.
DR InterPro: IPR006163; PP_bind.
DR InterPro: IPR006162; Pantne_attach.
DR InterPro: IPR01031; Thioesterase.
DR Pfam: PF00698; Acyl_transf. 1.
DR Pfam: PF0109; ketoacyl-synth. 1.
DR Pfam: PF02801; ketoacyl-synth. C_1.
DR Pfam: PF00550; PP-binding; 2.
DR Pfam: PF00975; Thioesterase; 1.
DR PROSITE: PS00075; ACP DOMAIN; 2.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; FALSE NEG.
KW Transferase; Acyltransferase; Phosphopantetheine; Repeat;
KW Multifunctional enzyme.
FT DOMAIN 383 814
FT DOMAIN 884 1209
FT DOMAIN 1706 1777
FT DOMAIN 1830 1901
FT DOMAIN ? 2181
FT ACT_SITE 552 552
FT ACT_SITE 978 978
FT ACT_SITE 1738 1738
FT BINDING 1862 1862
FT ACT_SITE 2028 2028
FT ACT_SITE 2181 AA; 238831 MW; 5A3B5712AA9AD942 CRC64;
SQ
Query Match 15.0%; Score 1083; DB 1; Length 2181;
Best Local Similarity 25.9%; Pred. No. 2.3e-49;
Matches 402; Conservative 235; Mismatches 637; Indels 280; Gaps 53;
QY 7 BBAADP-----IIVGSCRLPGGVLDLGSFTLGSRTVGRVPAERDAAAMPDP 60
Db 370 ERYSHRPGSDRCKLAIIVNSGHRFPAP-STDSFMFLTKGLDYCKEVPARRVDVTHYDP 428
QY 61 DPADGKTPVTRASFSLDVACFDASFFGISPREALRMDPAHRLLEVCMALENNAAIAPS 120

```

Db 429 SGKARKKATRMGCWLDFAGEFDPFRFSSISPKKAPQWDPQRMALMSTYEMERGIVDP 488
 QY 121 ALVGTG---TGVTIGIGPSEYEALPQATASAEIDAH---GGLGTPMSGARISYALG 173
 Db 489 TTPSTORNRIGVGHGVTNDW---METNTAONIDVYFTTGKRGFLP---GRINCFEE 540
 QY 174 LRPCCAVPTAYSSSLVAVHLACOSLSRGCESTALAGVSLMSPSTLWLSKTRMLARD 233
 Db 541 FSPSPISNDTACSSSLAHLACNSLMRGCDCTAVAGTMTIPTDGHSLDLDGFLSKRT 600
 QY 234 GRCKAFSAEADGFGREGCAVVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNG 293
 Db 601 GNCKAFDDADGVCRAEGVGTVPFKRLLEDALAEENDPLATILDIKTHSMSMSMTFRPK 660
 QY 294 SSGEIVLKRLLAAGCASSVGVENAGCTTGTIGDPIEIOALNAVY---LGDVATPRL 350
 Db 661 PAQIDNMSALLSTAGISPLDLSYIEMHGTGTQVDAVEMESVLSLPADETFPRDXPLX 720
 QY 351 IGSVKTNLGHPYASGITGLKLVLSLHOGQIPAHLLAQAALNP--RI--SMGDL---RLT 403
 Db 721 VSGAKANIGHGEVSGVTSILIKVLLMKNDTIPEHC---GIRKGSRLNRYVPLPARNVH 777
 QY 404 VTRACTPMDMNTPRRAGVSSFCMSGTNAHVLEAPPAATCTPAPERPAELLVLSART 463
 Db 778 IAFEPKPMPTDTPRRVLINNFSAGGNTAVLVEDAPV-----PELGEIAGSFG 826
 QY 464 SAlDADAARLRDLLETTPSQCLDVAFSLATTSSAMHRLAV--AATSRBELRALDLAAQ 522
 Db 827 EAAGNOSSEL-----SYTTARRMHPHRSVITGANTMELLRVESIAR 871
 QY 523 GQSPCAVRSIADSRGKLAFLFTGOGAQTLAGMGRLGYDVWSAFR---EAFDLCVRLFNQ 579
 Db 872 GH---GVNR---PATRKVIYIACSGOSQYGTGMQWLYNYPFRSLDEFPDLAASXGF 925
 QY 580 ELDRPLREVMABPA--SVDAALLDQTAFTQPALFTFEVALAALMWSGVPELVAGHST 637
 Db 926 P---SELEVTYSKRPVGDSEMDL---FVQLALVLSLEMLGLNLGSGFLKPAVIGHSL 979
 QY 638 GELVACVAGVPSLEDVAVFLVARGRLMQLPRAG--AMVSIAPEDVAAVAPHASV 695
 Db 980 GEYAAIYISGVLSAADTLVYVGMKAKLQERCORGHAMLAVRASPTLCEVLA--ESNC 1037
 QY 696 SIAAVNAPDOVVIAGAGOPVHAIAAAMAARGARTKALVHSAFSPMLAPMLBAFGRVAE 755
 Db 1038 EVACHNGPNVTVLSGLPKETMNLQNSLSAIGIKTLKLFATHSAGVQIILEFKNVAR 1097
 QY 756 SVSYRRPSIVLVNSGKACTDE--VSSPGYVWRHAREVRA---DGVKALHAAGACTFY 811
 Db 1098 GVTFHKFOIPLSLPLVAVKVIDEKGTVPVYLARRCHREPRVKNVSVLEHARDQHIITDRIV 1157
 QY 812 -EYCGPSTLLGLVPACMPDARPALILASSRAGDEPAVLEALGLMAVGLVSMAGL--- 867
 Db 1158 IDVGPRKALMGMTKTLIDKDTSSALPLTGPSLVDWKSILTILGLYSRGIDIMVVAHER 1217
 QY 868 FPGGGRVPLPTYPWQRERYWIDTKADDAAGDBRABAGAGDEVEBGAVRGSGRSARL 927
 Db 1218 FSGAKVTELPSTGYMDLKVFIPIKSGWCHL-----RHEIRGCAIPKGETATSQY 1268
 QY 928 DHPPEPSGRREKVEAAGDRPRLFEIDPG---VLHLVLVYTERRAVGLGEVEIAVDA 963
 Db 1269 QLSRDE-----QVAAKRPSKODESKEAVPEIATTVHARVVEKTEPIGAT--LIVET 1319
 QY 984 GLSFNDV-QIALGVNPPDLP-----GKPNPILLGSCAGRIYAVGSG-VNGLV-VGQPV 1035
 Db 1320 DISRPDNOIAQGLVDGIPLCTPSVYADIALHVGYSNMRLRSHDAGVVDVADVMV 1379
 QY 1036 I---ALSAGAFATVTTSAALVLPPOALSIEAAMPVAVLTPAVYLDRIARLQPERV 1092
 Db 1380 IDKALIPHGKSPQRLKTLTLMTP--PKAAATTSAAKIKFATYPRADGLD-----1427
 QY 1093 LHAATGCGVGLAAYVMAQHVGAEBVHATA-----GTPEKRAVYLSIG-----VR 1135

Db 1428 -----TEHATCTVEFTSEAOKSLQKVPXEYOEIRIKLGBGLRGQCFTR 1471
 QY 1136 YVSD-----SRSDRFVADVRAMTQGEVDVIVANSLSGELID-KSPNLRSHGRFVELGK 1188
 Db 1472 YTTKSGYKLMSSMASFHRDYKILN-----HILLNEADNEAVSTMDSAAKSECTF---AA 1523
 QY 1189 RDCYAD--NQLG-----LRPFLRNLFSVLV-----DLR 1214
 Db 1524 HAYVDAIIVQGVGFAMNANDNTDIOQEVFVNHGWTSSFQYQPLVKGKTYEYVVRMTEDBK 1583
 QY 1215 GMLLEKPAVY---RALLEELGL-IAAGVFTPEPPLATPIAVVADAFRMAQAOLHGLV 1270
 Db 1584 GDLVHEDTIVLGDAAVAFPKGLSVLSHLSSTQLTSSVYVRVPRRGRLMVLQASDKAA 1643
 QY 1271 LTLGDEP-VQIRLPTAGAGPSTGDRDLDRLASAAPARAAALAEAFRTQVSQVLRTP 1329
 Db 1644 RLHGNQAVKTOAFQRA-----LKQAPQSSFTPPHASKVAYSSAISPATGKV 1693
 QY 1330 I-----KYGAELFTRLGMDSLMAVELRNRIEASLK 1360
 Db 1694 VVARDLREGDDKFKAVLSVISESGVALGELTADTNFADIGIDSLSSVIGSRLLEDIG 1753
 QY 1361 LKLSL-----TFLSTSPNIALAQNLLDALATLALSLERVAENL 1399
 Db 1754 LELGAEFSLFDICPTVRSUKTLILSGS--AVSVNDKDELPEGQEAETAPAEQL 1804
 RESULT 13
 PAS_HUMAN
 ID PAS_HUMAN STANDARD; PRT; 2504 AA.
 AC P49327;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Fatty acid synthase (EC 2.3.1.85) [Includes: EC 2.3.1.38; EC 2.3.1.39; EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.10; EC 3.1.2.14].
 GN FASN OR FAS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=96004605; PubMed=7567999;
 RA Jayaramar A., Tai M.-H., Huang W.-Y., Al-Feel W., Hsu M., Abu-Elheiga L., Chitrata S.S., Wakil S.U.;
 RT "Human fatty acid synthase: properties and molecular cloning.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:8695-8699 (1995).
 RN [2]
 RP SEQUENCE OF 753-758 AND 1285-1297.
 RX MEDLINE=94294385; PubMed=8022791;
 RA Khatjida F.P., Jemner K., Wood F.D., Hennigar R.A., Jacobs L.B., Dick J.D., Pasternack G.R.;
 RT "Fatty acid synthesis: a potential selective target for antineoplastic therapy.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6379-6383 (1994).
 CC - FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THIS MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN ACYL CARRIER PROTEIN.
 CC - CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a long-chain fatty acid + (N+1) CoA + N CO(2) + 2N NADP(+).
 CC - CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA + acetyl-[acyl-carrier protein].
 CC - CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA + malonyl-[acyl-carrier protein].
 CC - CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) + [acyl-carrier protein].
 CC - CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] + NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
 CC - CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein]


```

Db 1137 OLCKGLVEALETKVQOGLKMWVPDWTGPRSPQCEHPRLLSAACRLQNGNLQLEL 1196
      886 -----RYMTDKADD-----AARGD--RARPG 905
Db 1197 AQVLAQERPKLPEDLLSGLLSDPALKACIDTAENNPSTLMKKEVLAAGHLYSRIPG 1256
      906 -----AGHDEVEEGAVRG----- 919
Db 1257 LLSPHPLLOLSTYATDHPQALEAAQOELOOHVAAQOMPRADAPASLSADLLVNCNA 1316
      920 ----GDRSARLD-----HP-----PPESGR----- 936
Db 1317 VAALGPASALSNMVAALREGFLLTLTRGHPRDIVAFLTSTEPQYGGIISQDAMES 1376
      937 ---REKVEAG-----DRPRLEID----- 953
Db 1377 LFSRVSRLVLKKSFGYATFLCRPPTPODSPIFLVFDTSFRWESLKGILADEDSR 1436
      954 -----EPG-----VLDH 960
Db 1437 PVLKKAINCATSGVGLVNCLEPRPGTAVCVLLSNLSTSHVEVDPGSAELQKVLQCD 1496
      961 LVLRVTERRAPGL-----GE 975
Db 1497 LVNVVVRDGMGVFRHFLLEDKPEEPAHAFVSTLTRGDLSTIRWCSLIRHAQPTCPGA 1556
      976 VELAVDAAGLSFNDVOLALG-MWPDRLRGK-PNPRLLLGEGCGRIYAVEGNGVLVQO 1033
      1557 QLCCTVYASINFRDIMLATEKLSPDALPGMTSODSLKMEFSGR-DASGRVWGLV-- 1612
      1034 PVIALSGAPATVHTTSALVLPRLPQALSAIEAAMPVATLTMVALDRJAPLOPGERVL 1093
      1613 -----PAKGLATSVLSPLDMVPSWMTLEBAASVAVYSTAYALVAGRRPGETTL 1667
      1094 IHAATGCVGLAAVQMAQVGAEVHATAGTEPKRAYLES-----LGVVSDSRSDRPAVDY 1149
      1668 IHSGSGGVQAAIALIALSLCGRVFTTGALEKRAYLQARPPQDSTSFANSRPTSFQHV 1727
      1150 RAMTGGCGVUVVNSLSCGLIDKSFNLRSRGRFVELGKDCADNOLGLRPFRLNLSFS 1209
      1728 LMTGGGVDLVINSLSAEELQASVRCFTHGRFLELGKDLSONHPLGMAILFKVTFH 1787
      1210 LVDLRGMMLEPRPARVALLEELGLIAGVFPPTPIATLPIARVADAFRSMAOQHLGKL 1269
      1788 GVLDAFNNSSADMREVALVERAIRDGVVRPKCVFHGAYVEDAFRMAQKIHGV 1847
      1270 VLTGDBEVOIRIPTHAGAGPSTGDRDLRLASA-----APAR-----AAALEAFLRTQ 1320
      1848 VVQVLAEEPAVL-----KGAKP-----KLMSAISKTECPAHKSYIIAGIGLGF 1893
      1321 VSGVLTPEIKVGAEE-ALFTPLGM-DSLMAVELRNRLEASLKLKJSTLST 1370
      1894 LAQWL-----IQRGVQKVLTSRSGIRGTQAKOVRRWRROGLQOVSTSISS 1942

```

RESULT 14

```

PKSK_BACSU STANDARD; PRT; 4447 AA.
AC P40803;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksk (PKS).
OS PKSK.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / P81424;
RX MEDLINE=95219083; PubMed=7704258;
  Albertini A.M., Caramori T., Scoffone F., Scotti C., Galizzi A.;

```

```

RT "Sequence around the 159 degree region of the Bacillus subtilis
RL genome: the pksx locus spans 33.6 kb.";
RT Microbiology 141:299-309(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=93843377;
  Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
  Azevedo V., Berteiro L.G., Bessieres P., Bolotin A., Borchert S.,
  Brouillet S., Bruchet C.V., Caldwell B., Capiano V., Carter N.M.,
  Choi S.K., Codani J.U., Conterton I.F., Cummings N.J., Daniel R.A.,
  Deniot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
  Eutican K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
  RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
  RA Ghim S.Y., Glaser P., Goffeau A., Golighly E.J., Grandi G.,
  RA Giuseppe G., Guy B.J., Hega K., Haiech J., Harwood C.R., Henaut A.,
  RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
  RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
  RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
  RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
  RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
  RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
  RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
  RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
  RA Prescan E., Pujić P., Putrelle P., Rapoport G., Rey M., Reynolds S.,
  RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sataie Y.,
  RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
  RA Sekiuchi J., Sekowska A., Serot S.J., Serot P., Shin B.S., Soldo B.,
  RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
  RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
  RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
  RA Viari A., Wambuit R., Wedler H., Wedler H., Weitzenecker T.,
  RA Winere P., Wipat A., Yamamoto H., Yamane K., Yatsunoto K., Yata K.,
  RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Darchin A.,
  RT "The complete genome sequence of the Gram-positive bacterium Bacillus
  RT subtilis.";
  Nature 390:249-256(1997).
  CC - FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
  CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
  CC SECONDARY METABOLISM.
  CC - COFACTORS: CONTRAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
  CC (POTENTIAL).
  CC - SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
  CC FAMILY.
  CC - SIMILARITY: Contains 5 acyl carrier domains.
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-eb.ch/announce/
  CC or send an email to license@isb-eb.ch).
  CC
  CC EMBL; U11039; AA85144.1; -
  CC EMBL; 299112; CAB13590.1; -
  CC EMBL; 299113; CAB13601.1; -
  CC PIR; A69679; A69679.
  CC HSSP; P14687; IAMU.
  CC Subtilat; BG10930; pksk.
  CC InterPro; IPR000873; AMP-bind.
  CC InterPro; IPR001242; Condensation.
  CC InterPro; IPR000794; Ketoacyl-synt.
  CC InterPro; IPR006163; PP_bind.
  CC InterPro; IPR006162; Ppantne_attach.
  CC Pfam; PF00501; AMP-binding; 1.
  CC Pfam; PF00668; Condensation; 1.
  CC Pfam; PF00109; ketoacyl-synt; 3.
  CC Pfam; PF02801; ketoacyl-synt_C; 3.
  CC Pfam; PF00550; PP-binding; 5.
  CC PRINTS; PR00154; AMPBINDING.
  CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.

```

DR PROSITE; PS00455; AMP BINDING; 1.
 DR PROSITE; PS00606; B KETOACYL SYNTHASE; 2.
 DR PROSITE; PS00075; ACP DOMAIN; 5.
 KM Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
 KM Phosphoenolpyruvate; Multifunctional enzyme; Repeat; Ligase;
 KM Complete proteome.
 FT DOMAIN 1 68
 FT DOMAIN 1063 1130 ACYL CARRIER (ACP) 1.
 FT DOMAIN 2516 2589 ACYL CARRIER (ACP) 2.
 FT DOMAIN 2618 2687 ACYL CARRIER (ACP) 3.
 FT DOMAIN 3868 3937 ACYL CARRIER (ACP) 4.
 FT BINDING 31 31 ACYL CARRIER (ACP) 5.
 FT BINDING 1093 1093 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT BINDING 2552 2552 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT BINDING 2650 2650 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT ACT SITE 2915 2915 PHOSPHOPANTHETHEINE (POTENTIAL).
 FT BINDING 3900 3900 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 FT ACT SITE 4147 4147 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
 SQ SEQUENCE 4447 AA; 496058 MW; 9C6DB4AC0A9C057 CRC64;

Query Match 14.1%; Score 1018; DB 1; Length 4447;
 Best Local Similarity 26.3%; Pred. No. 1,4e-45;
 Matches 391; Conservative 203; Mismatches 519; Indels 376; Gaps 52;

QY 2 ADPIERAAE-DPIAVGASCRLLPGGVLDLGGFWLLBGRDTPVGRVPERMDAAAF-D 59
 DB 2733 ABRNKKQADFPVALVIGISGRFP-GAMIDEFWKNLEGGKDSITEVPRKDRMDREHYGN 2791
 QY 60 PPPDAGKTPVTRASTLSVACPDASFEGISPREALRMPARHLLLEVCEALENNAIP 119
 DB 2792 PPTDV-NKTDIKWGGFIDVAFDPLFFGISPREADYVDPQOQLMTYWKLEDEGCSF 2850
 QY 120 SALVGETGVFICIGPSEYEALPQATASAEIDAHGGLTM-PSVGAGRIAYALGRPC 178
 DB 2851 QSLSTGICIFIGTGTGYKDLFHR-ANLPIEGHAATGMIPSVGNMSTFLLHGS 2908
 QY 179 VAVDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMLSPSTLVMLSKTRALARDRCCKA 238
 DB 2909 EBEVETACSSSLVAIHRAVTAQNGCEMALAGVNTILTEAHISYSAKMLSTORCKT 2968
 QY 239 FSAEADFGRRGCAVYVILKLSGARADDRILAVIRGALNHGASGLITVPNGSSQRT 298
 DB 2969 FSADANGVYRGGVGMVWLKLEDERDGNHIVGVIRGAENHGGANLITSPNPAQAD 3028
 QY 299 VIKRALADAGCAASSGVYEAHGTGTLGDIPIEQALNAVY---GLGRVAPLL----- 350
 DB 3029 LILVRAVROADIPSTVYTIENAGTGTGLDPIEINGLKAFKELSMRSESQDPVDFHC 3088
 QY 351 -IGSVKTLGHEBYASGITGLKVLVSLQHQI PAHLHAQALNPRIISMGDLRLTVTRAT 409
 DB 3089 GIGSVKSNIGHLELAAGISGLIKVLLQMGKTLVSLHCETLNPYLQGLDSFPYIQEKO 3148
 QY 410 PVP-----DMN-TERRAGVSFGMSGTNAHVLEE-APATCTPPAPERPAELVLISART 462
 DB 3149 EKSVYVDRGNELPRAGISSFGIGGVNNAIYIEETMPANSEHTTEOP-NVIVLSAKN 3207
 QY 463 ASALDAQARLDHL--ETYPQCGDVAFSLATRSAMEHLAVATSRREGIRALDA 520
 DB 3208 KSRLLDRASQLEVRANKKTYTODDLRIYTLQVGEEDERACVAGTMOELEKQLRF 3267
 QY 521 AOGQTSFGAVRSIADSSRGKLAFLFTGCAQTLGMRGLVYWSAFREAFDLCVLFNOE 580
 DB 3268 VDGKETDEFFRGGSHRNKETQITIFADDMALAL-----DAMIRKTKAKLA----- 3315
 QY 581 LDRPLREVMWAPASVDALLDQTAFTOPALFTF-----EVALALMRSGVEPELV 632
 DB 3316 -----DLMVGVSIQNTL--YGETKPRLLISLSPYPAKHVWPAHESEERKKEIV 3366
 QY 633 AGHSIGELVACVAGVSLDAVFLVAAGRLMQLPAGAVNSIEAR-AVAAVA-- 689
 DB 3367 --NALEDRAACFLTKWMSLP-----IGSAVPGRTVAIILCCQGTALDALEVSSY 3414
 QY 690 -BHAASVSIAAVNAPO-----VVIAGAG-----QPVHAIAMAAARGA 727

DB 3415 FNNHLLDIVSRLE-NDQSDIDWKEFDGLVDVIGCGMDDEGRUDMIETWQRLVEFGHKEGL 3473
 QY 728 R-----TKL-----HVSIAFSPILMAPLEAFGRVAESVSRPBI 764
 DB 3474 RLLCVTKGLESPQNTSVWAGASRAGLYRMQCESSHLSIRMDA-----EEVTDIRR--- 3526
 QY 765 VLVSNLSGKACTDESSPGYWRHAREVRFADGVKALHAAGATFEVEGPKSTLLGLVP 824
 DB 3527 -----LAKLADVEYSYSDAE-----VCYRDGR-----YQAFKAPETGATBOS 3569
 QY 825 ACPDAPALLASSRAGDEPATVLEALGVLAVAGVLSWAGLPFGSGGRVLPPTYPMQR 884
 DB 3570 AVFP-----KDH-----VLLINGTRGIGLCA----- 3592
 QY 885 EHYWIDTKADDAARDRAAPAGHVEVEGGAVRGDRRSALDPPRESGR----- 937
 DB 3593 -RHFAB-----CYGVKKLVLTGREQL-----PPREWARFKTSNTS 3627
 QY 938 --EKVEAAGDRPFRLIDEFVL-----DHLVLRTERAPAGLGEVIAVDAAGL 985
 DB 3628 LAEKIQAV-----RELAKGVQVEMLSLTSDAQVQGTQHIGRTIGIGVHICAGL 3681
 QY 986 SFNDVQALGVPPDLPCKENPPLLGGECAGRIYAVGEGVGLVVGQPVIALSAGAFAT 1045
 DB 3682 TMDTLAIFIRKTSDDIQVLEPK-----VSGLT-----TLVR 3713
 QY 1046 HTTSAALVLRPQALSI--EAAAMPVAYLTAWALRIARLOGERVLHATGCVGL 1103
 DB 3714 HVCNPELOFPVLFFSSVAIIPELSGQADYMANSYMYPFAAHQKAPIT----- 3764
 QY 1104 AAVQNA--OHVGAETHATAGTEPKAYLESIGVRVVSRSRDFVADVRAWTGEGVDV 1161
 DB 3765 -SVQPMNKEG-----MGEVTHQAYRDS--GLSTINSBGRFIDQVSKKFG--PVY 3813
 QY 1162 LNSLSEGLIDKSFNLSRSHGRFVELGKRD CYADNQLGRPFLRNLSFSLVLDRCMMLER 1221
 DB 3814 LPAMA-----NQTNMEPEL-----LMKRX 3833
 QY 1222 ARVRLLEELGLIAGVFTPPITLPIARVADFRSMAQOHGKVLTLTGDEPVOIR 1281
 DB 3834 PHEGLQCE-----ALQSPPA-----RDIIEBDEVSK----- 3860
 QY 1282 IPTHAGAPSTGDRDLRLRLASAPARAALAEFLRQVSOVLTPEIKVGAELFTRL 1341
 DB 3861 -----CGGLSET-----QSLVILDFBELRIDDEPFDGLFQDY 3896
 QY 1342 GMDSLMAVELNRRIEASLKLSTFLSTSPNIALLAQNLDAATALS 1390
 DB 3897 GVDSTILAQVLRIRKLEALDPEILVEYPTIQRFADWLIGSYVERLS 3945

RESULT 15
 FAS CHICK
 ID FAS CHICK STANDARD; PRT; 2511 AA.
 AC P12376;
 DT 01-OCT-1998 (Rel. 12, Created)
 DT 01-FEB-1998 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Fatty acid synthase (EC 2.3.1.85) [includes: EC 2.3.1.38; EC 2.3.1.39;
 DE EC 2.3.1.41; EC 1.1.1.100; EC 4.2.1.61; EC 1.3.1.1.10; EC 3.1.2.14].
 GN FASN OR FAS.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=90311;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.
 RC STRAIN=white leghorn; TISSUE=liver;
 RX MEDLINE=95031085; PubMed=7944406;
 RA Huang W.-Y., Chitkala S.S., Wakil S.J.;
 RT "Amino-terminal blocking group and sequence of the animal fatty acid

```

RT synthase."
RL Arch. Biochem. Biophys. 314:45-49(1994).
RN [2]
RP SEQUENCE OF 75-1775 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=89282777; PubMed=2734291;
RA Holzer K.P., Liu W., Hammes G.G.;
RT "Molecular cloning and sequencing of chicken liver fatty acid
  synthase cDNA."
RL J. Biol. Chem. 264:3750-3757(1989).
RN [4]
RP SEQUENCE OF 1752-2512 FROM N.A.
RX MEDLINE=8830436; PubMed=2842766;
RA Yuan Z., Liu W., Hammes G.G.;
RT "Molecular cloning and sequencing of DNA complementary to chicken
  liver fatty acid synthase mRNA."
RL Proc. Natl. Acad. Sci. U.S.A. 85:6328-6331(1988).
RN [5]
RP SEQUENCE OF 2202-2512 FROM N.A.
RX MEDLINE=89088152; PubMed=3207710;
RA Kasturi R., Chirala S.S., Pazirandeh M., Wakil S.J.;
RT "Characterization of a genomic and cDNA clone coding for the
  fatty acid synthase gene."
RL Biochemistry 27:7778-7785(1988).
RN [6]
RP SEQUENCE OF 2121-2209.
RX MEDLINE=89192401; PubMed=2648999;
RA Huang W.-Y., Scoops J.K., Wakil S.J.;
RT "Complete amino acid sequence of chicken liver acyl carrier protein
  derived from the fatty acid synthase."
RL Arch. Biochem. Biophys. 210:92-98(1989).
RN [7]
RP SEQUENCE OF 2209-2508.
RX STRAIN=White Leghorn;
RC MEDLINE=89088151; PubMed=3207709;
RA Yang C.-Y., Huang W.-Y., Chirala S.S., Wakil S.J.;
RT "Complete amino acid sequence of the thioesterase domain of chicken
  liver fatty acid synthase."
RL Biochemistry 27:7773-7777(1988).
RN [8]
RP SEQUENCE OF 667-674 AND 1698-1709.
RX MEDLINE=89323081; PubMed=2751995;
RA Chang S.I., Hammes G.G.;
RT "Amino acid sequences of pyridoxal 5'-phosphate binding sites and
  fluorescence resonance energy transfer in chicken liver fatty acid
  synthase."
RL Biochemistry 28:3781-3788(1989).
RN [9]
RP FUNCTION: FATTY ACID SYNTHETASE CATALYZES THE FORMATION OF LONG-
  CHAIN FATTY ACIDS FROM ACETYL-COA, MALONYL-COA AND NADPH. THIS
  MULTIFUNCTIONAL PROTEIN HAS 7 CATALYTIC ACTIVITIES AND AN ACYL
  CARRIER PROTEIN.
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + N malonyl-CoA + 2N NADPH = a
CC long-chain fatty acid + (N+1) COA + N CO(2) + 2N NADP(+).
CC -1- CATALYTIC ACTIVITY: Acetyl-CoA + [acyl-carrier protein] = CoA +
CC acetyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + malonyl-[acyl-
CC carrier protein] = 3-oxoacyl-[acyl-carrier protein] + CO(2) +
CC [acyl-carrier protein].
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxyacyl-[acyl-carrier protein] +
CC NADP(+) = 3-oxoacyl-[acyl-carrier protein] + NADPH.
CC -1- CATALYTIC ACTIVITY: (3R)-3-hydroxypalmitoyl-[acyl-carrier protein]
CC = 2-hexadecenoyl-[acyl-carrier protein] + H(2)O.

```

```

CC -1- CATALYTIC ACTIVITY: Acyl-[acyl-carrier protein] + NADP(+) = trans-
CC 2,3-dehydroacyl-[acyl-carrier protein] + NADPH.
CC -1- CATALYTIC ACTIVITY: Oleoyl-[acyl-carrier protein] + H(2)O = [acyl-
CC carrier protein] + oleate.
CC -1- SUBUNIT: HOMODIMER, WHICH IS ARRANGED IN A HEAD TO TAIL FASHION.
CC Event=Alternative splicing; Named isoforms=2;
CC Name=2;
CC IsoId=P12276-1; Sequence=Displayed;
CC Name=1;
CC IsoId=P12276-2; Sequence=VSP_000149;
CC -1- CAUTION: REF.3 AND REF.5 SEQUENCES DIFFER FROM THAT SHOWN FROM
CC POSITION 2351 ONWARDS DUE TO A FRAMESHIFT.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U04485; AAA86389.1; -.
CC EMBL; J03860; AAA48767.1; -.
CC EMBL; J02839; AAA82106.1; ALT_SEQ.
CC PIR; S57248; XYCHA.
CC InterPro: IPR001227; Ac transferase.
CC InterPro: IPR002085; Adh_zn_family.
CC InterPro: IPR000794; Ketoacyl-synt.
CC InterPro: IPR006162; pp_bind.
CC InterPro: IPR001031; Thioesterase.
CC Pfam; PF00698; Acyl_transf_1.
CC Pfam; PF00107; Adh_Zinc_N_1.
CC Pfam; PF00109; ketoacyl-synt_1.
CC Pfam; PF02801; ketoacyl-synt_C_1.
CC Pfam; PF00550; pp-binding_1.
CC Pfam; PF00975; Thioesterase_1.
CC PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
CC PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
CC PROSITE; PS50075; ACP_DOMAIN; 1.
CC Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
CC Transferrase; Hydrolyase; Oxidoreductase; Lyase; NADP; Acetylation;
CC Alternative splicing; Pyridoxal phosphate.
CC KM
CC INIT MET 0 0
CC DOMAIN 1 7411
CC DOMAIN 427 815
CC DOMAIN 1638 1866
CC DOMAIN 1867 2119
CC DOMAIN 2124 2180
CC DOMAIN 2209 2511
CC MOD_RES 1 160
CC ACT_SITE 160 160
CC ACT_SITE 579 579
CC ACT_SITE 877 877
CC NP_BIND 1674 1691
CC BINDING 1707 1707
CC NP_BIND 1888 1903
CC BINDING 2157 2157
CC ACT_SITE 2308 2308
CC ACT_SITE 2481 2481
CC VARSPIC 2348 2348
CC CONFLICT 77 78
CC CONFLICT 116 116
CC CONFLICT 675 675
CC CONFLICT 1169 1169
CC CONFLICT 1178 1178
CC CONFLICT 1191 1191
CC CONFLICT 1198 1198
CC CONFLICT 1286 1287
CC FTID=VSP_000149.
CC QI->PV (IN REF. 2).
CC L->A (IN REF. 2).
CC L->S (IN REF. 2).
CC K->N (IN REF. 2).
CC R->T (IN REF. 2).
CC R->H (IN REF. 2).
CC P->L (IN REF. 2).
CC DN->ND (IN REF. 2).

```



```

FT CONFLICT 1372 1372 K -> E (IN REF. 2).
FT CONFLICT 1533 1533 C -> Y (IN REF. 2).
FT CONFLICT 1577 1577 W -> R (IN REF. 2).
FT CONFLICT 1685 1696 OAAIATLSMGC -> ASHCHRLHGLA (IN REF. 2;
AAA48767).
FT CONFLICT 1732 1732 Q -> E (IN REF. 2).
FT CONFLICT 1745 1745 S -> N (IN REF. 2).
SQ SEQUENCE 2511 AA; 274648 MW; 622039DAC8315D3F CRC64;

Query Match 13.8%; Score 997.5; DB 1; Length 2511;
Beet Local Similarity 21.7%; Pred. No. 8.6e-45;
Matches 425; Conservative 186; Mismatches 564; Indels 783; Gaps 43;

QY 12 DEIAIVGASCRPLPGVIDLSGFMTLESGRDTYGRVPAERMDAAMFDDPDPAKPTPV 71
DB 1 EDVVIAGIAGKLPESF-NIQEFEMNLNGVDWYTE-DDRRMKPGIYGLP-----K 48
QY 72 RASFSDVACPDASFPGISPREALMDPAHRLLEYCEWALENAAIAPSAIVGTETGPT 131
DB 49 RRGKLDIKKFPASFPGVHPKQAHMTDPOLRLLEVSYEAILDGGINPFLALRGTDGTGVV 108
QY 132 GIGSEYEALPOATASAEIDAAGLGTMPSVGAGRISVALGRGPCVAVDTAVGSSIVA 191
DB 109 GASGBEALFALSO--DPEELIGISMTGCPRAMLANRISFYDPTGSLTIDTACSSSLMA 166
QY 192 VHLACOSLRSGECSTALAGVSLMLSPSTLWLSKTRALARDGRCFAFSAEADGFGREG 251
DB 167 LENAIAKAIHGGCSALVGVGNVILKPTNSVOFMKLMSPDGCACAFVSGNGYCRSA 226
QY 252 CAVVVLKRLSGARADGDRILAVIRGSAINHGD-ASSGLTPNGSGOEIVLKRALADAGA 310
DB 227 VVVVLLTKKSMK---RYVATIVAGSNTDGFKEGQVTFPSGEMOOLGSLYREGCIK 282
QY 311 ASSVGVYEHAGGTTLGDPIEIOAL-NAYYGLGRDVAATPLIGSVKTNIGHEVYASGITG 369
DB 283 PGDVEVEHAGGTGTGKGDPOEVNGIVNVEFCOCERE--PLLTGSTSNNGHEBPASGLAA 339
QY 370 LKVVLSLOHQOIPAHILHAQALNPRISW--GDLRLTVTRARTPMDMTTPRAG--VS 423
DB 340 LAKVLSLEHGLMAPLHNDPNPDIPALHDSGLKVVCK-----PTVKGGLVSVIN 390
QY 424 SFGMSGTNAHVLEB---APAATCTPPAPERPAELVLVSARTASALDAQARLRDHE 478
DB 391 SFGFGSGNAHVILRPEKEKCOPOETCNLP-----RLVQVCGRTQRAVEILIEESRKHG 444
QY 479 TYP-SQCGDVAFLATITSSAMEHRLAVALTSREGRLALDAAOCOTSPGAVRSIAD-- 535
DB 445 CSPFLSLSDIS--AVPVSSMPYR-----GYTLVGTESDITETIQ 481
QY 536 ---SSRGKLAFLFTGCGAQTLLGMRGLYDVMSAFREAFDLCVRLFNOELDRPLREVMMAE 592
DB 482 QVQAGRGPLMTYICSGMGTOMKMGSL-----MKDLDFRQSLIRS-DELYKST 528
QY 593 PASVDAALL-----DOTAFTOPALFTEYVALAALMRSGVPELIVAGSHIGELVAACV 645
DB 529 GLKVSDDLINADENFTDDTVHAFVGLAAIQIADIVLKAAGLPDGLIHSHVIGELACGVA 588
QY 646 AGVFSLIEDAVFLVAARGRLMO--ALPAGAMVSIAPENDVAAVAAPHASVSIANVAP 703
DB 569 DNSLSHEEVLAAYWRGRCVKEAKLPFGG-MAAVGLTWCEKQORCP--NVVPACHNSE 644
QY 704 DQVVIAGAGQPVHAIAAAMAARGARTKAL-HVSHAPEPLMAPMLEAF----- 750
DB 645 DTVVYSGPLDSVSEFTYTKLKGQGVRAKEVRAGVAFHSTYMASIAPALLSALKVIYPRK 704
QY 751 GRVAVSVYRPSIVLVNLSGCACTDEVSSPGYVWRHAREVVRPADGVKALHAAAGAGTF 810
DB 705 PRSAWISTSIPIESQOSDLARN-----SSAEYHVNNLVNPLYFIEGK-K-HIDENAVV 756
QY 811 VEVGRKSTLLGLVPACMPDARPAALLASSAQRDEPVTVEALGGLMAVGLVSMAGLPPS 870
DB 757 VEIAHPALLQALIRTLRTKPTCTLLPLMKDKHKNLEFPLTQTKIHLTGINVLGNLFP 816

```

```

QY 871 GGRVYPL-----PTYWQREERYWIDTKAADDAARGDRAPGA-----G 907
DB 817 VEYVPVGTPLISPYIKKDHQDMVPAEDFPSSGKSASAVYNIDVSPDSPHYLVG 876
QY 908 H----- 908
DB 877 HCIDGRVLVPATGYLVLAARLTARSLGNVMEQTAVMFEEVTIHOATILPKKGSQLEVRI 936
QY 909 ----- 908
DB 937 MPASHFEVSGNGNLAVSGKISLENDALKNFHNQADPOSOQANTASGLMEDVQEL 996
QY 909 ----- 908
DB 997 HLRGYNPTFGVLECNSEGSAGKILWNGMWVTFPLDTHLILVLAETGRSLRPTIRSR 1056
QY 909 ----- 915
DB 1057 VYIDVHLHQBVOYQODNVEAFDVVVDRLDLSLKAGVQINGLASVAPRQOERISPTL 1116
QY 916 -----AVRGDRRSARLDHP 930
DB 1117 EKFSFVPIESDCLSSSTQILAHYLECKGLIQKLOAKMALGVKLVHGLETKGAASGP 1176
QY 931 PPSGRRE-----KVEAGDRPFRLP-----IDEP--GYLDHLVLRV----- 965
DB 1177 PAQKGLQHILREICGELNPNHSELBOIVTQEKMHLDDPFLNGLDLSSELKTCLDVAK 1236
QY 966 -----TERRAR- 972
DB 1237 ENTTSRHKIVEALAGSRLFSRVQSIINTPOLLDYIATDCTPBTETLSDNETELHDAGI 1296
QY 973 ----- 972
DB 1297 SFGQDPSSLPSGNLTNADLAVCNCSITVLTGNTAEIISNLAAYKEGFEVLHTLKEET 1356
QY 973 LGEV----- 976
DB 1357 LGEIVSLTSPDLQOKHSFLSOAQWELFSKASLNLVAMKRSFSGSVIFLCRRQSPAKAP 1416
QY 977 -----EIANDA----- 982
DB 1417 ILLPVDDTHYKVVDSLKEILDSBPMLTATNGNSGILGMVNCRLBEAGHRIKCVF 1476
QY 983 ----- 982
DB 1477 VSNLSPSSTVPATSLSLSEMOKIIERDLVNAVYRDCKWGSFRHLPLQQAPOELTECAV 1536
QY 983 -----AGLSFNDVOLALG-MVPDDLPGK-P 1005
DB 1537 NVLTGDSLRLRWIVSLRHHQTTNPVQLOCKVYVASLNFMDIMATGKLSPDALPGMW 1596
QY 1006 NPPLLGECAGRIYAVEGVNGLVGQPVITALSAGAPATHVTSAAVLVPRQALSAIE 1065
DB 1597 LQOQMLGMEFSGRDLA--GRRVNGL-----LPKAGLATVDDCDKREPLMEVPEMWTLEE 1647
QY 1066 AAAMPVATLTAMYALDRLARLOPGERVYLIIHAATGVGLAAVQMAONVAGVATAGTPEK 1125
DB 1648 AASVPVVAATAYVALVVRGKMGKGSVLHSGSGGVQALAIATLSMCKGRFATGSAK 1707
QY 1126 RAYLES-----LGVRYVSDRSRDRFVADVRAVMTGGEVDVNLNLSGELIDKSFNLRSHG 1181
DB 1708 REYLQARFPQDANSFASSTRTTPOQHILRVITNGGSLVNLNLSAEKLGASRLCLAHG 1767
QY 1182 RFEVLEKDCYADNOLGLRPLRLNLSFSLVDLRGMMLERPARVRLLELLGLIAGVFT 1241
DB 1768 RFLIEIGKEDLNNNSOLGVALFLKVVAFHGIILDSIFEEGNOBWEVVELTTGIGIDGVVK 1827
QY 1242 PPIATLPIARVADAFRMAQOHLGKLVTLTGDEYQ 1279
DB 1828 PLRTTFGKEEVEAARFMAQOKHIGKWKIKIQEEBK 1865

```

Sat Oct 4 14:14:34 2003

us-10-014-717-2.rsp

Page 24

Search completed: October 2, 2003, 17:28:40
Job time : 54.2576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 26.2576 Seconds
(without alignments)
2544.975 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210

Sequence: 1 VADRIERAEADPPAIVGAS.....GVQNDPVSSGADQDWELIAL 1421

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2107.5	29.2	3567	1	ERY2_SACER
2	1923.5	26.7	3491	1	ERY1_SACER
3	1889	26.2	3519	1	OL56_STRAT
4	1875	26.0	3172	1	ERY3_SACER
5	1852.5	25.7	2110	1	MCAS_MYCBO
6	1592.5	22.1	1538	1	PPSA_MYCTU
7	1583	22.0	1876	1	PPSA_MYCTU
8	1180	16.4	1986	1	WA_EMENT
9	1177	16.3	1774	1	MSAS_PENPA
10	1173.5	16.3	2109	1	PKS1_ASPPA
11	1100	15.3	2505	1	FAS_FAT
12	1083	15.0	2181	1	STCA_EMENT
13	1039	14.4	2504	1	FAS_HUMAN
14	1018	14.1	4447	1	PKSK_BACSU
15	997.5	13.8	2511	1	FAS_CHICK
16	940	13.0	4427	1	PKSL_BACSU
17	882.5	12.2	4273	1	PKSM_BACSU
18	612.5	8.5	352	1	FAS_ANGAN
19	392.5	5.4	317	1	FABD_BACSU
20	363	5.0	401	1	NOB4_RHIME
21	354.5	4.9	402	1	NODE_RHIME
22	344	4.8	308	1	FABD_SALTY
23	336	4.7	403	1	NODE_RHIV
24	332	4.6	308	1	FABD_ECOLI
25	329.5	4.5	467	1	KASA_STRCO
26	328	4.5	312	1	FABD_HAEIN
27	325.5	4.5	402	1	NODE_RHIS3
28	324.5	4.5	401	1	NODE_RHIL1
29	321.5	4.5	426	1	KASI_STRGA
30	316.5	4.4	300	1	VAT1_HUMAN
31	315	4.4	412	1	FABF_ECOLI
32	308	4.3	414	1	FABF_VIBHA
33	304.5	4.2	293	1	FABD_SYNY3

34	303.5	4.2	413	1	FABF_VIBCH
35	303	4.2	838	1	FAS_MOUSE
36	302	4.2	420	1	KASI_STRCM
37	298.5	4.1	327	1	QOR_ECOLI
38	298	4.1	421	1	KASI_STRVN
39	295.5	4.1	327	1	QOR_SALTY
40	293	4.1	421	1	FABF_RHIME
41	289.5	4.0	422	1	KASI_STRHA
42	285.5	4.0	416	1	FABF_SYNY3
43	284.5	3.9	423	1	KASI_STRCO
44	282	3.9	379	1	VAT1_TORCA
45	278	3.9	325	1	QOR_PSEAE

ALIGNMENTS

RESULT 1
ID ERY2_SACER STANDARD; PRT; 3567 AA.
AC 003132; 054096;
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 3 and 4 (EC 2.3.1.94) (ORF 2) (6-
DE deoxyerythronolide B synthase II) (DEBS 2).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
OC Pseudonocardiales; Pseudonocardaceae; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=91220065; PubMed=2024119;
RX Donadio S., Staever M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide
RT biosynthesis.";
RL Science 252:675-679 (1991).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2338;
RX MEDLINE=92155230; PubMed=1740151;
RA Bevil D.J., Cortes J., Haydock S.F., Leadley P.F.;
RT "6-deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT Cloning of the structural gene, sequence analysis and inferred domain
RT structure of the multifunctional enzyme.";
RL Eur. J. Biochem. 204:39-49(1992).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC deoxyerythronolide B.
CC -1- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHEINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC RESPECTIVELY.
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: Contains 2 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.fab-sib.ch/announce/> or send an email to license@fab-sib.ch).

CC EMBL; M63677; AAA26494.1; -;
 DR EMBL; X62569; CAA44448.1; -;
 DR PIR; S23070; S23070.
 DR InterPro; IPR001227; Ac transferase..
 DR InterPro; IPR002085; Adh zn family.
 DR InterPro; IPR000794; Ketoacyl-synt.
 DR InterPro; IPR006163; Pp_bind.
 DR InterPro; IPR006162; Ppantne_attach.
 DR Pfam; PF00698; Acyl_transferf; 2.
 DR Pfam; PF00107; ADH_zinc_N; 1.
 DR Pfam; PF02801; ketoacyl-synt; 2.
 DR Pfam; PF00550; Pp-binding_2; 2.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 2.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
 DR TRANSFERASE; Acyltransferase; 2.
 KM Phosphopantetheine; Multifunctional enzyme.
 FT DOMAIN 1 1484
 FT DOMAIN 1485 3567
 FT DOMAIN 27 488
 FT DOMAIN 559 884
 FT DOMAIN 1130 1301
 FT DOMAIN 1397 1467
 FT DOMAIN 1485 1943
 FT DOMAIN 2013 2336
 FT DOMAIN 2383 3066
 FT DOMAIN 3139 3322
 FT DOMAIN 3415 3485
 FT ACT_SITE 202 202
 FT ACT_SITE 651 651
 FT BINDING 1430 1430
 FT ACT_SITE 1661 1661
 FT ACT_SITE 2105 2105
 FT NP_BIND 2961 2978
 FT NP_BIND 3142 3157
 FT BINDING 3448 3448
 FT BINDING 438 438
 FT CONFLICT 480 480
 FT CONFLICT 1241 1241
 FT CONFLICT 2664 2664
 SQ SEQUENCE 3567 AA; 374413 MW; EEC284F4738AA0C0 CRC64;

Query Match 29.2%; Score 2107.5; DB 1; Length 3567;
 Best Local Similarity 32.5%; Pred. No. 4.5e-103;
 Matches 610; Conservative 179; Mismatches 530; Indels 559; Gaps 45;

QY 2 ADRIPIERAAE-DPLAIVGACRLPGGVIDLSGFMTLL-EGSRDVGVPNER-WDAAMF 58
 DB 1478 AVIVVGAADESEPIAIVGICRPPGIGSPBQLRWVIAEGANLTG-FPADRGVDIGRLY 1536
 QY 59 DDPDAAGKTPVTRASFSDVACFDASFFGISPREALRMPAHRLLEVCMEALENAIA 118
 DB 1537 HPPPDNGCTSYVDKGGFTLDADEDFPGFTTREPALAMPQGRIMLETMEWEVERGID 1596
 QY 119 PSALVGTGTGVTGIGPSEYEALPQATASAEIDAHGIGLGTMPBVGAGRI SYALGLRGC 178
 DB 1597 PDALRGDTGVFVGMNGSQSYQLL--AGEARVVDGYGLGNSASVLSGRITATFGWEGPA 1654
 QY 179 VAVDTAASSSLVAVHLACQLRSGECTALAGCISLMSSTLVMLSKTALARDGCKA 238
 DB 1655 LTVDTACSSSLVGHLMQALRSGESLALAGVTVWSDYTFVDFSTQGLASDGRCKA 1714
 QY 239 FSAEADFGGEGGCAVVVLKRLSGARADGRILAVINGSAINHDGASSGLTVPGSSOEI 298
 DB 1715 FSAEADFGGEGGCAVVVLKRLSGARADGRILAVINGSAINHDGASSGLTVPGSSOEI 298
 QY 299 VLKRALADAGCAASSVGVYEAHGTITLGDPIETIQALNAVYGLORDVATPLLTGSVXTNL 358
 DB 299 VLKRALADAGCAASSVGVYEAHGTITLGDPIETIQALNAVYGLORDVATPLLTGSVXTNL 358

DB 1775 VIRQALAAAGVPADVDVVEAHGTETELDPIEAGALITATGGDRD--RPLRGSVXTNI 1832
 QY 359 GPEVYASGITGLIKVYLSTOHGOIPAHILHAQNLNPRISMGDLRLVTRARTPMDWNTPR 418
 DB 1833 GHTQAAAGAAAGYIKVYLAKRHGKLPRLSLHADELSPIIDBESGAVEVLRREVPAPERR 1892
 QY 419 RAGVSSFGNSGTNAHVLEEAR-----AATCTPPAPERPAELLVLSARTASALDAQARL 473
 DB 1893 RAGVSSFGVSGTNAHYIVEAPAEQEAARTENGPL-----FVLSGRSEAVVAQAARAL 1946
 QY 474 RDHLETPYPOCCLDVAFSLATTRSGAMEHRLAVALTSREGRLAALDAAGQTSFGVRSI 533
 DB 1947 AEHLRPTPELGLDAMWTALATGARFDVAAYVGGDRACVCELDALAEGRPADVAVP 2006
 QY 534 AOSRSGKLAFLFTGGAQTILGMRGLYDVWASAFREAFDLCVRLFNQELRPLREVMAB 593
 DB 2007 TSAFR-KPVLVFPGQAGQWGAHARDLLESSEVPFAEBSMCAELSTHTWKLLDVRRG 2065
 QY 594 ASYDAALLDQTAFTQPAFTFEYALALWRSWGBEELVAGHSIGELVACVAVGSLE 653
 DB 2066 GPDHERVD---VLQPVLSIMVSLAEIWRAGVTPAAVVGSHQGEIAAHVAGALSLE 2122
 QY 654 AVFLVARGRLMOALPAGAMVSTIAPADVAANAAPHASVSIANAAPDOVIVTACQ 713
 DB 2123 AAKVVALRSGVIRELDQGMVSVGASRDELETVLRMDGRVAVAAVNGPSTSVAGPTA 2182
 QY 714 PVYALIAAAMARGARKTALHVSAPSLMAPPLAEFGRVAVESVYRPSIVLSNLSGK 773
 DB 2183 ELDEFFAEAEARMKRRRIAVRYASHPEVAREDLAELELITTVAGRSVPLHSTVGE 2242
 QY 774 ACTDEVSSPGYVWRHAREVRFADGVKALHAAGAGTFVEVGPKSTLL-----GLV 823
 DB 2243 VIDTSAMDASVWYRNLRPLVFEQAVRGLVEQCFDFVEVSPHVLVMAVEETAHAGAE 2302
 QY 824 PACMPARPALLASSAGDEPATVLEAL--GGLAVGGLSVAGLP--SGRRKPLPT 879
 DB 2303 VTCVP-----TLREQSGPHEFLRLRAHVHG--ADLRPAVAGGRPAELPT 2349
 QY 880 YPMQREYV--IDTKADDAARGDRA----- 903
 DB 2350 YPEHQRFPFRPRPVDVSHLVGRGAEHPLLAADVPGHGAVFTGRISTDEQPLAEH 2409
 QY 904 -----PG-----AGHD----- 909
 DB 2410 VVGRTLVGSLVDLALAAGEDVGLPVLBEVLQRPVLVAGALLRMSVGAPEDESGRR 2469
 QY 910 ----- 909
 DB 2470 TIDVHAEDVADLADQMSQAHATGLAOGVAAGPRDEQWPEEDAVRIPLDHIDGLAQ 2529
 QY 910 -----EV-----EEG-----GAV----- 917
 DB 2530 GRYEGSFQALRAAMKDSVYAEVSIADDEBGIAPHVLLDAVQTLISGLALGERGGK 2589
 QY 918 -----RGDRRSARL----- 930
 DB 2590 LPEAMTVTLHSGATSVRVATPACADAMALRVTPAGHLVATVDSLVRSTGEKWEQ 2649
 QY 931 PPESGRE----- 947
 DB 2650 EPRGEGEELHALDMGRLAERGSTGRVVAADASDLVLRSGEPEPPAVLVRYEPEDDPR 2709
 QY 948 ----- 947
 DB 2710 AAARHGVLAALVRMLEOEBELPGATLVATISGAVTVSDSDSVPEPAAAMGVRCQ 2769
 QY 948 -----FRLEID-EEGVL-----DHLVLRV-----TERAPG 972
 DB 2770 AESPDFFVLLDTPAEFGMLPAVPDNPQALRGDDVFPRLSPAPSAULTLPACTQRLVP 2829
 QY 973 -----LGEVELAVDAAGISFNDVOLALGMVDDDLPGKRNPELLG 1012
 DB 2830 DGAIDSVAFEPADVEQPLRAGEVRVDVTRATGVNFRDVLALGMYPQKAD-----MG 2881

```

QY 1013 GCACGIVAVGCVGLVVGOPVIALSAGAFATHVTTSAALVLRPQALSAIEAAMPYA 1072
DB 2882 TEAGVAVTAVGVDVAFAGDVRVLGFGAFAPAIATDHRLLARVDGSDADAAVPIA 2941
QY 1073 YLTAMVADRIRARLOGERVLTHATGVGLAAVQAVGAEVHATCTPEKRAYLESU 1132
DB 2942 YTTAHVALDLHLAGRAGOSVLTHAAAGVGMVAVALARRAGAELVLTAG-PAHGTLRAL 3000
QY 1133 GV-RVPSRSDRFADVAMTGGCVVUNLSGELIDSPNLSHGPRVLEGRD 1190
DB 3001 GUDDEHIASSKETGFRKRTERTGSGVUNLSLTGELDESADLLAEDGVFVENGKTD 3060
QY 1191 CVADNOLGURPFLRNLSFSLVLDLRCGMLEPPA-----RVRLLEELLGLIAGVF 1240
DB 3061 -----LRDAG-----DPRG-----RYAPFDLGEAGDRLGELILEVVGILGAGSL 3100
QY 1241 TPEPIATLPIANVADAFRSMQAQHLKGLVLTG--DPEVOIRIPTHAGSPGTDRDL 1297
DB 3101 DRLPVSAMELGSAPALQHMRSGRHVGLVLTQAPAVDPDGVLI-----TGCTGT 3151
QY 1298 LDRILASAPARAALAEFLRQVSVLTPREIKVGAELFTRLGMSLMAVELRRRIA 1357
DB 3152 LGRLL-----LARHLVTEHGVRL-----LVSRGADAGSDLELRAEIE- 3190
QY 1358 SIKLKLTFTLSTPNIALLAQNLDALATL-----SLERVAEN-LR 1400
DB 3191 DGGASMEIACPTADBDALSA--LDDGLRPLTGVVHAAGVADGLVTSIDEPAVEQVLR 3248
QY 1401 AGVQNDVSSGADQDWEI 1418
DB 3249 AKV-----DAANWL 3257

RESULT 2
ERYL_SACER STANDARD; PRT; 3491 AA.
ID ERYL_SACER
AC 003131;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-deoxyerythronolide B synthase 1) (DEBS 1).
GN ERYA.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91220065; PubMed=2024119;
RA Donadio S., Steaver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT "Modular organization of genes required for complex polyketide biosynthesis."
RL Science 252:675-679 (1991).
RN [2]
RP SEQUENCE OF 3474-3491 FROM N.A.
RX MEDLINE=93231529; PubMed=8386127;
RA Donadio S., Steaver M.J.;
RT "Is1136, an insertion element in the erythromycin gene cluster of Saccharopolyspora erythraea."
RL Gene 126:147-151 (1993).
CC -1- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-deoxyerythronolide B.
CC -1- CORFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC -1- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN BIOSYNTHESIS.
CC -1- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3 ORF CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5, AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,

```

```

CC CC
CC -1- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR)
CC DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC OF THE FULL-LENGTH CHAIN.
CC -1- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC -1- SIMILARITY: Contains 3 acyl carrier domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63676; AAA26493.2; -.
CC EMBL; L07626; AAA26504.1; -.
CC PIR; T43231; T43231.
CC HSSP; P25715; TMLA.
CC InterPro; IPR001227; Ac transferase.
CC InterPro; IPR002198; ADH short.
CC InterPro; IPR000794; Ketoacyl-synt.
CC InterPro; IPR006163; Pp bind.
CC InterPro; IPR006162; Pantine attach.
CC Pfam; PF00698; Acyl_transf_3.
CC Pfam; PF00106; adh_short_1.
CC Pfam; PF00109; ketoacyl-synt; 2.
CC Pfam; PF02801; ketoacyl-synt_C; 2.
CC Pfam; PF00550; pp-binding; 3.
CC PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
CC PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
CC PROSITE; PS00705; ACP_DOMAIN; 3.
CC TRANSFERASE; Acyltransferase; Antibiotic biosynthesis; NADP, Repeat;
CC KW Phosphopantetheine; Multifunctional enzyme.
CC -----
CC FT DOMAIN 1 1972
CC FT DOMAIN 1 1979 3491
CC FT DOMAIN 1 375
CC FT DOMAIN 414 484
CC FT DOMAIN 503 961
CC FT DOMAIN 1030 1356
CC FT DOMAIN 1611 1794
CC FT DOMAIN 1888 1958
CC FT DOMAIN 1979 2441
CC FT DOMAIN 2507 2854
CC FT DOMAIN 3055 3237
CC FT DOMAIN 3334 3404
CC FT ACT_SITE 145 145
CC FT BINDING 447 447
CC FT ACT_SITE 677 677
CC FT ACT_SITE 1128 1128
CC FT NP_BIND 1614 1660
CC FT BINDING 1921 1921
CC FT ACT_SITE 2148 2148
CC FT ACT_SITE 2598 2598
CC FT NP_BIND 3058 3104
CC FT BINDING 3367 3367
CC SEQ SEQUENCE 3491 AA; 365022 MW; 682BFC3C90FA8C4 CRC64;

Query Match 26.7%; Score 1925.5; DB 1; Length 3491;
Best Local Similarity 35.6%; Pred. No. 1.7e-93;
Matches 547; Conservative 184; Mismatches 601; Indels 203; Gaps 39;

QY 4 RPIERAEDPIAIVASGRLPGVVIDLSGFWTLGSRDVTGVRAER-MDAAMFDDPD 62
DB 497 RINEAAGPEVAVVMAACRLPGVGSTPREPFWLLSEGRDAVAGLPTDRQWDDLSFHPDP 556
QY 63 DARGKTPVTRASFSDVACFDASFGICSPRELRNDPAHRLLEVCWELEMAALAPSL 122
DB 557 TRSGTAHORGGGFLTEATAFDPAFFGMSPRELAVDPOORLMLLSWEVLERAGIPPTSL 616

```

QY	123	IGTEGVYIGIGPSPEYEAALLPOATASADIDAGHIGTMPEYVGRASISYALDGRGCYAVD	182
Db	617	QASPTGVYVGLIPOEYGRPL--ABEGGEVEGYLMTGTTTSTASGRATITLLGEGPAISVD	674
QY	183	TAYSSSLVAVHLAQCSSLRSGECSTALAGVSLMSPSTLYWLKXTRALARDGRCAFSAE	242
Db	675	TACSSSLVAVHLAQCSSLRSGESSLAMAQGVYVMPPTGMLVDPSRNNISLAPGRCAKAFSAG	734
QY	243	ADGREGGECGCAVVLKRLSGARADGDRLLAVITRGAITHDGASSLTYVPNNSOGEYVLAKR	302
Db	735	ANGFEMAGAGMLTLERLSDARBNHGPLYAVLRGTAVNSDDASNGLSAPNRADQVRYLQO	794
QY	303	ALADAGCAASSGYEAGHTGTGTDPIEIOALNANVYLGGRDVATPTLLIGSVKTNLGHPE	362
Db	795	ALAESGLCPADIDAEAHGTGRJSDPIEABALFEAY--GKDREOPHLHVSXKSNLGTQO	852
QY	363	YASGITGLKVLSLKHGOIPAHHLAQAOLNBRISGWDLRLVTRBATPPEMPTPRRAGV	422
Db	853	AAAGVAGYIKVILAMRAGTLPTRLTHASRSKSEIDSSGAILLDEPEPWPAGARRRAGV	912
QY	423	SSFGMSGTNAVHVEEAP-----AATCTPPAPERBAELVLSAFTASALDAQAARL	473
Db	913	SSFGISGTNAHAIIEEAPQVVEGEVEAGDVAP-----VWLSASASAGLRAQAARL	964
QY	474	RDHLETYSQCLGDVAFSLATTTSRAMEHRLVAASREGLRALDAADAAQOTSQCAVRSI	533
Db	965	AAHLLEHNGOOPRDIAYSLATGRALPRHAPAPVDESALRLVDGLATGNAQDAAGVGT-	1023
QY	534	ADSSRGKALPFTQCGAQTGLGKGRSLYVWMSAFREAPDLCVRLFNOELDRPLREYMAEP	593
Db	1024	-SRAGQRAVVFPPGGQMGWAGNVDLDTSPVFAALALECADLEPHLDFEVIFFLREA	1082
QY	594	A--SYDAL--LDQTAFTPPALPTFEYALALMRSGVPELVAHSGISGELVAAQVAGVS	650
Db	1083	ARREDDALSTERVDVQVWFVAWVSLASMMRAGVYPAVITGHSGEIIAAACVAGALS	1142
QY	651	LEDNAFVLAAGRLMQLPAGGAMVSTIAREADVAAVAAPHAASVSTIANVAPROQVYAG	710
Db	1143	LDDAARVALRSRVLATPMPGKNMVASIAAPGEVYARIGDR--VEINAAGVPRSSVAVAG	1199
QY	711	AGQPHAIATAAMAAARGARTKALHVSASHSPILMAPMLBAFGRVAESVYRPSSTV-LVSN	769
Db	1200	DSDELDRLVASCTTECIATKRLANDYASHSHSVETIRALALAELEDPHRLPGVPPFST	1259
QY	770	LSGKACTDEVSSPGYVWHAREEVYRFADGVXALHAAGACTFVEVGPXSTLLGLVPACMPD	829
Db	1260	VTGRWTQDELDAGYWMRLRRTYVAFADVABALAEQVRTFLEVSAPHILTAIEIIDG	1319
QY	830	ARPALLA--SSRAGDEPARYLEALGGLMAYGGLVSNAGL-PPSGGRVVPPLPTPWQER	886
Db	1320	SGADISAIHSIRRGGSLADPGEALSRFAAGVAVDWESYHIGTGARVRPLPTYPFORER	1379
QY	887	YWIDTKADDAARGDRPAGAGHDEVEEGAVR-----GGDRRSARLD-----	928
Db	1380	WMLEK--PVARRS-----TEVDEVSALRYRIEMPTGAGBARLDGTMLVAKTAGT	1429
QY	929	HPPESSGRREKVEAAGDRPFRLEIDEPGVLDHLVLTERRAPGLGEVE-----IAYDAA	983
Db	1430	ADETSTAREALESAGARRELVNDCARGDELA-----ERLSVGEVAGVLSILLAVDEA	1484
QY	984	GLSFNDVOLALGMVDDLP-----GKNPPLLIGGECAGRIYAVGSGVNGLV--V	1031
Db	1485	--EPEEAPLALASLDLTSLVOAMVSAELGCPMLVTESAVATGPFERVNRMAAGALMGV	1542
QY	1032	GOPTALASAGAPATHVTTSALVTLVRPQALSAIE-----AAAMPVAVLTVMYALDR	1082
Db	1543	GVAVIALENPAYWGLVDVPAQSVAEIAHLAHLAAVNSGAGGEQOLALRADGVYGRRWTRAA	1602
QY	1083	IA---RLQPERVLLHAATGVGGLAAVQMAOCHVGAEVNA-----TACTPEKRAYLE	1130
Db	1603	PATDEMKRPTGVVLTGTGCGVGGQIAMWLARRGAP-HLLVLSRSGRPADAGELVABLE	1661
QY	1131	SLGVRY-----VSDSRSDRFADVAVRAMTGGEGVDVVL-----NSLSGEL	1169

```

Db      1662  ALGARTVAACDVTDRRS-----VRLLGGIGDDVLSAVFPAATLDDGTDTLTGER 1715
Oy      1170  IDKSFNLIRSHGRVVEIGKDCY-ADNOLGRPPLRLSL-----VDLR 1214
Db      1716  IERA-----SRKAVIGARNHETRELDLTFVLFSSFASFGAGLGGYAPGNAYLD 1768
Oy      1215  GMLER-----PAVR-----LLEELLGLIA-----AGFTPP-----1244
Db      1769  GLAQQRSDDGLPAIVAVMGTWAGSGMAE---GAVADFRHGVTEMPEPACRALONALD 1825
Oy      1245  ---IATLPiARVADAFPSMAQAQHLGLVLTGDPPEVOIRIPTHAGAPSTGRDLLDLRL 1301
Db      1826  RAECPIVIDRWKRFLLATAGQPRFLPBEIDAR---RAAPAPAEPRVG-----AL 1876
Oy      1302  ASAAPAAAPAAALEAFRTQVSQVL-RPEIRKGAELFTLGMDSLMAVELRNRIEASLK 1360
Db      1877  ASLAPEREELFELVSRHAAVIGHASARVPADQAFAEIGVDSLSALELRNLGTAATG 1936
Oy      1361  LKSTPLSTSPNALLAQNULDLALALSLERA 1395
Db      1937  VRLPTTVFDHPDVRTLAAHLMAELGATGAEQAA 1971

RESULT 3
O556_STRAT STANDARD; PRI: 3519 AA.
ID_O556_STRAT
Dt      01-NOV-1997 (Rel. 35, Created)
Dt      01-NOV-1997 (Rel. 35, Last sequence update)
Dt      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Oleandomycin polyketide synthase, modules 5 and 6.
GN      ORFB.
OS      Streptomyces antibioticus.
OC      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC      Streptomycinae; Streptomycetaceae; Streptomyces.
OX      NCBI_TaxID=1890;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=94150470; PubMed=8107683;
RA      Swan D.G., Rodriguez A.M., Vilches C., Mendez C., Salas J.A.;
RT      "Characterisation of a Streptomyces antibioticus gene encoding a type
RL      I polyketide synthase which has an unusual coding sequence.";
RL      Mol. Gen. Genet. 242:358-362(1994).
CC      -I- FUNCTION: MAY BE INVOLVED IN THE BIOSYNTHESIS OF THE OLEANDOMYCIN
CC      LACTONE RING.
CC      -I- COFACTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC      -I- SIMILARITY: Contains 2 acyl carrier domains.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC      or send an email to license@ebi.ac.uk).
CC      -----
CC      EMBL: L09654; AAA19695.1; -.
DR      PIR: S43048; S43048.
DR      HSSP: P25715; IMLA.
DR      InterPro: IPR001227; Ac transferase.
DR      InterPro: IPR000794; Ketoacyl-synt.
DR      InterPro: IPR006163; Pp_bind.
DR      InterPro: IPR006162; Pantane attach.
DR      InterPro: IPR001031; Thioesterase.
DR      Pfam: PF00698; Acyl_transf. 2.
DR      Pfam: PF00109; ketoacyl-synt. 2.
DR      Pfam: PF02801; ketoacyl-synt_C; 2.
DR      Pfam: PF00550; pp-binding; 2.
DR      Pfam: PF00975; Thioesterase; 1.
DR      PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.
DR      PROSITE: PS00606; B_KETOACYL_SYNTHASE; 2.
DR      PROSITE: PS50075; ACP_DOMAIN; 2.

```


KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine; Multifunctional enzyme; Repeat.
FT DOMAIN 1 3519
FT DOMAIN 2 501
FT DOMAIN 3 569
FT DOMAIN 4 880
FT DOMAIN 5 1200
FT DOMAIN 6 1487
FT DOMAIN 7 1561
FT DOMAIN 8 1686
FT DOMAIN 9 2220
FT DOMAIN 10 2541
FT DOMAIN 11 3038
FT DOMAIN 12 3215
FT DOMAIN 13 3270
FT DOMAIN 14 3519
FT ACT SITE 210
FT ACT SITE 660
FT NP BIND 1203
FT BINDING 1524
FT ACT SITE 1859
FT ACT SITE 2311
FT NP BIND 2859
FT BINDING 3178
SQ SEQUENCE 3519 AA; 368561 MW; 41AE78AAAE61F86 CRC64;
Query Match 26.2%; Score 1889; DB 1; Length 3519;
Best Local Similarity 34.0%; Pred. No. 1.4e-91;
Matches 541; Conservative 191; Mismatches 595; Indels 264; Gaps 44;
9 AAEDIIATVGCACRLPGGVITDLSGFWTLESGSRDTPYGRVPAER-NDAAAMPDDDPARCK 67
1664 ASREAIATATMNSCRGGGIDSPEDIMRFLAEGRDAAVAGLPEDRGMDLALYHPDDEPCT 1743
QY 68 TPVTRASPLSDVACEASFFGISPBEALMDPAHRLLEVCWEALENNAIAPASLVGET 127
DB 1744 TVRGCAFHYDAAGDAGFFGISPREALMDPQRLLETSELEPERADIDYTVRGTAI 1803
QY 128 GVFTIGTIGSEY--EALDPAATASABIDAHGGLGTMPSVAGARISYALGRCPCVAVDTA 184
DB 1804 GFIFAGHGYGPDPRAPESVAGYLL-----TGTAVALSGRISYTFGLGPAVTVDTA 1858
QY 185 YSSSLVAHVLACOSLRSGECSTALAGVSLMSPSTLVLSKTRALARPGRCKARSAMD 244
DB 1859 CSSSLVLAHLAAVQALRGECSTALAGVAVKSTPDAVEFSSROQGARGRCKARPAAMAD 1918
QY 245 GFGREGCAVVVVKELSGARADGDRILAVIRGSAINHDGASGLTVPNSSOEIVLKRAL 304
DB 1919 GKGWEGSVSLLEHLELSDARRIGHVLAIVRGSAVNOGDSNGLAENPSPGQRIYIRAL 1978
QY 305 ADAGCAASSVGYVEAHGTGTTLGDPIEIQALNAVYGLGRDVAATPLLIGSVKTNLGHPEYA 364
DB 1979 ADAGLAPADVDVVEAHGTGTRLDPIEAQALLATYGGGRAGRPVWLGSVKSNIQHTQA 2038
QY 365 SGITGLLVVLSLQHGQIPAHILHAQALNPRISMGDLRLTVTAARIPM-PDMTTPRAGVS 423
DB 2039 AGVAAVMKQVNLGKGVVPKTLHVDSPHVDMSAGAELLTEERPEWEAEARLRAGIS 2098
QY 424 SFGMGSTAAHVLEAPAPATCTPPAPERPAELL-----VLSARTASALDAQAARLR 474
DB 2099 AFGVSTAHAVIVEAPR-----EPPEPEPTRVAAAGDLVVPVWVSGRVDGALREDAARLA 2154
QY 475 DHLETTYPSQCGDVAFSLATTRSAMEHRLAVALTSREGRLAALDAAAGQTSPPGAVRSIA 534
DB 2155 AHVSTGTAGVV-DVGMSTLVAATRSVEHRAVMVGTDLDSMAGSLAGPAAGVVGVSVA 2213
QY 535 DSSRGKLAFLFTGGCAQTLGKMGRLYDVMSAFREAFDLCVRLFENELDRPLREVMABPA 594
DB 2214 PABGRVVFVFPGGQSQVGMMAAGLLDACPVFAEVAECBAALVDVTGMSLVEIQGR-- 2271
QY 595 SYDALLDOTAFTOPALFTFEYALALMRSWGVPEBELVAGHSIGELVAAVCAVGSLEDA 654
DB 2272 --DAVTLGRVDVQPALMAVMVSLARTKRYVEVEPAAYVGHSGEIAAACVAGGSLADG 2329
QY 655 VFLVARGELMOLPAGGAMVSIETAPADVAAVAAPHAASVIAAVNAPDOVVIAGAP 714

DB 2330 ARVVLRSRARIARIAGGGMVSLSLPAGRVRTMLDTYGGRVSAVAANGPSSVTVSGDYQA 2389
QY 715 VVAIAAAMAAGARTKALVSHAFSPPLMAPLEAFGRVAVSVSRPSIVLVNSLSCGA 774
DB 2390 LDELLAGGEREVRKRRVPDVASHSAQMDQIRDLLENALADITQDSSVFPFSTVTADW 2449
QY 775 CTDEVSSPGYVVRHAREVVRPADGVKALHAAGAGTFVEVPGKSTLLGLVPACMP--DARP 832
DB 2450 LDTTALDAGYFNTLRRETVRQEAIVEGLVAQGMGAFVSCSPHPVLVPGIEQLDALDQNA 2509
QY 833 ALLASSRARGDEPA--TVLEALGLMAVGLVSYNA-----GLFPGSGRRVPLFTYMQER 886
DB 2510 AVLGSIR--RDEGLDRLLTSLAEAFVQGVPMVDMTHAFEGVTP--RTVDLPTYPEFORQ 2564
QY 887 YVIDRKADDAARG--DRRAPGAGHDEVEGAVR-----GQPRRSARLDHPPEESRREV 940
DB 2565 FMLDSSPASSANGVDGEADAMITWDVEREDSVAABEIGTIDEALHTVLPALSSRRRRV 2624
QY 941 E-----AAGDRPFLLEIDE-----PGVLDTLVLRTERRAPGLGE 975
DB 2625 EHRRRLQDMRYVEMKFPFPAALDEVLGGMFLVPRGLADDGVAVVAATVARG--GE 2680
QY 976 VEI-----AVDAAGLSFNDVOLALGMP--DDLPGKPNPPLLIGGECA 1016
DB 2681 VSVVELDPTRPDRRAYAEAVAGRGVS-----GVVSFLSWDDRRHSEH-FVVPAGLAA 2731
QY 1017 GRIVA-----VGEQVN-GLVVGQPVYA--LSAGAATHTVTTA-----ALVLPKQ- 1059
DB 2732 SLVLAQALVDLGRVGEGRMLVTRDAVAVAPSDGAVIDPQAOVGFGRVLGEHPEL 2791
QY 1060 -----ALSAIEAAA-----MPVAVLYAWYALDRIRAL----- 1086
DB 2792 MGGLIDLVLEAPPEPGSTCDHTYADLATVVASAGEDDVAARGSGVW--VRLVAAVVDG 2849
QY 1087 -----QGEERYLIIHAATGCGVGLAAVQAQVHGAEVH-----ATAGTEPKRAYLESL 1132
DB 2850 GGGGWRPRGTVLVYTGGLGGLGAHTARMLVGGAD-HVLLVSRGGSAPGAGDLVLEBGL 2908
QY 1133 GVRVYSDRSRPFAD--VRAWTG--GEGVDVYVNSLS----- 1166
DB 2909 GGARVSVRACD--VADRVALLSLSDGEPYTAFFHAAGVPOSTPLAEISVQEAADVMAA 2966
QY 1167 -----GELID-----KSFNLRSHGRFVELGKRDCA----- 1193
DB 2967 KVAAGVNLGELVDPGLEAFVLFSSNAGWMSGGCAVYAAANAFDLAALVRRGGLPAT 3026
QY 1194 -----DNQGLRPFLRNLSFSLVDLRGMLEPRPARVALLLELLGLIAAGVFTPPP 1244
DB 3027 SVAMGMWAGEGMAVSGAARELS-----RRGVRAAMDPERAVAVMAVAGREBAFV-----A 3077
QY 1245 IATLPIARVAAFMSMAQOHGKLVLTLDGPEVOIRI-----PTHAGAGPSTGDRDL 1298
DB 3078 VADVMEFVTGFAFSARPRPLISDL-----PEVRTALRNOBOQLHAPV-FEDBSAQL 3130
QY 1299 DRLASAPARAARAALEAFITQVSOVLRTPEIK-VGAELFTRLGMDSLMAVELRNRTIEA 1357
DB 3131 RRLMSLSPAGAEALVYKVRTEAAVAVLHGSAQDVPAERAKELGFSLTAVQLRNRLAA 3190
QY 1358 SLKULSTFLSTSPNIALLAQNLIDALATA 1388
DB 3191 ATGTRLPASAVFDHPHAAALARWLLAGMRHA 3221
RESULT 4
ERY3_SACER
ID ID ERY3_SACER STANDARD; PRT; 3172 AA.
AC 003133; 054097; 099270;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-deoxyerythronolide B synthase III) (DEBS 3).
GN ERYA.

QY 4 REPERAEDPIAIVGASCRLPGVJIDLSGFWTLLEGGSRDVTGRVPAERWDAAMPDPDD 63
 Db 31 REVERAGERPIAIVGACRFPDGDVDSPESEFWFVSGCGDAIAEAPDR-----GM-EPDSD 85
 QY 64 ABGKTPVTRASFLSDVACDASFFGGSPEALRMDPAHLLLEVCHEALENAALAPALV 123
 Db 86 A-----RLGMLAAGDADGAFPGISPREALAMPQOQIMLEISSEALERGHGHPVSIR 139
 QY 124 GTEYFVIGIGPSEYEALPOATASAEIDAHGGLGTMPSVGAGRISYALGSPCAVDT 183
 Db 140 GSATGVFTGVGTVDVGRPDE--APDEVGIVGTGTASSVAGSRVAVCLGEGSPMTVDT 197
 QY 184 AVSSSLVAVHLACQISRSECESTALAGVSLMSPSTVLMKTRALARDGRCKAFSAEA 243
 Db 198 ACSSGLTALHAMESIRDECCGALAGVTVMSSPAFTFPRSGGLAADGCKPFSKXA 257
 QY 244 DFGGEGGCAVVLKRLSGARADGRILAVIRGSAIINHGAASSGLTVPNGSQEYLKKA 303
 Db 258 DDFGLAEGAGVVLRLSARREGRVLAIVLAGSANOGASNGILAPSGPAQOQVIRRA 317
 QY 304 IADAGCAASVGVVEAHGTTLGDPPIEIOALNAVYGLRDVATPLLIGSVKTNIGHPEY 363
 Db 318 LENAGVRAADVVEAHGTITRLGDI EVHALISTYGAERDDPDLWISVKSNIQHTOA 377
 QY 364 ASGITGLKVLVLSLOHGOIPAHLHAQALNPRIISWMDLRLTVTRARTPMPDMNTPRAGVS 423
 Db 378 AAGVAGVMAVVALIRGEMPTLHPDESPQIEMDGAIVSVSQAASVAGERPRRAGVS 437
 QY 424 SFGMGSTNAHVLEAPAACTTPPAPERPAELIVTSARTASALDAQARLRDHLLETYSQ 483
 Db 438 SFGISGTNAHVLEAPADEPEPAPDSGPVPLVLSGRDEQAMRAQAGRLAHLAPEPAN 497
 QY 484 CIGDVAFSIATRSAMEHRLVAASRE---GIRALDLAAAGQTSRPAVRISIAOSSRK 540
 Db 498 SLRDGTFTLTASAMEHRAVVGDRDELALGIRAVADRRIADRTATGQ---GPSISPR 553
 QY 541 IAFLEFGGAGQTLGMRGLYDVWSAFREAFDLCVLRFNGLRPLEVW-NAEPASVDA 599
 Db 554 VAMVFPGGAGQMGMAARDLRESQVADSIRDCERLAHVHWSLUDLSGARP----- 607
 QY 600 LLDQTAFTOPALFTEYALALMRKSGVEPELVAGHSIGELVAACVAFSLIEDAVFLVA 659
 Db 608 -LDRVDVVOPALFAVWVSLAALMRSHGVPAVVGSHGEIAAAHVAGALTLEDAAKVA 666
 QY 660 ANGRIMQALPAGAWVSIAPADVAANAAPHAASVIAVANAP-DQVVIAGAGQVNAI 718
 Db 667 VASRVLRIRGGQGNASFGLTGEQAERIGRPAAGALVSIASVNGPRSVVVAAGEGDLDEL 726
 QY 719 AAMAMARGARTALHVSHPFSLMAMPLEAFGRVASEVSYSRPSIVLSNLSGKACTDE 778
 Db 727 IAECEAEHAKARRIPVDYASHSPQVESLREBELTELAGISPVSAVVALSTTTGQIDTA 786
 QY 779 VSSPGVWHAAREVVRADVKALHAAGAFTVEVGPKSTLLGLVPACMPDARA----- 833
 Db 787 TMDTYWYVYANLREOVYFQDARTQLAEAGFDFAVEVPHVPLTVGIEBATLDSALPDAGAC 846
 QY 834 LLASRAGDEPATVIEALGILMANVGLVSNAGLFPSSGGRRPPLPYPMQERERYWDTQA 893
 Db 847 VVGTRLRDGGIADPHTALGEAVYAGVEVDWSPAF-ADARPVELLPYPO--RYMPLIPT 903
 QY 894 DDAAR---GDRAPGAGHDEVEGCAVGRGDRRSARLDHP--PPE--SGRREVEAAGDR 946
 Db 904 GGRADDEDDMWYQVVMREAEWESASLAG---RVLLVTPGVPSELSDAIRSGLESGAT 960
 QY 947 PRLLEID-----EPGVLDHL---VLRVTERRA--PGLGEVEI--AVDAAG----- 984
 Db 961 VLTCDVESRTIGTALAEADDTALSTVGAAPVHGBAVPSPDIALVALGAAAGVADL 1020
 QY 985 --LSEFDVOLAIGMVPDDLPGKNPBLLLIGEGCAGIVAVGE-----GNGVIVGQPVIA 1037
 Db 1021 WVLTRAAVVADELVD--PAQ---AMVGG--LGVVIGIEQGRNGVLVDL----- 1065
 QY 1038 LSGAFATVHTTSAALVL-PRPOLSAIEAAMPAVVALTAWYALDRIRLQF-GERVLH 1095

Db 1066 ---DADAASIRSLAVALADPRGEEQVAIRADGIKVARIIVPAPARARTHPLEBLACTVLT 1122
 QY 1096 AATGVGLAAVOMAHVGAENHA-----TACTPEKRAYLESGLVRY-----VSDS 1140
 Db 1123 GGTGGIAGHLARWMLRSSAE-HLVLLGRGADAPASBLREBLTALGCVTTIACDVADR 1181
 QY 1141 RSDRFVADVRATVAGEGVVVLN---SLSGELDKSFNLRSRQRFVELGRDCYADNQ 1196
 Db 1182 ARLENVLAEEAAEBERTVSAMVHAAGVSTSTPLD-----LTAEFTEI--AD-- 1227
 QY 1197 LGLRPLNLNFSFLVDLRGMMLERPARVALLIEELGLIA---AGVFTPEPIATLPAR 1252
 Db 1228 -----VKVRGTV-----NDELCPDDDAFVLFSSNAGVWGSPELASVAAAN 1268
 QY 1253 V-ADAFRMAQA-----QHLGKVLVLTGP-----EVQIRI 1282
 Db 1269 AFLDFAPARAREGAPVTSIAGWLAGONMAGDEGEYLRSGCLAMPDRDRAVELHTTL 1328
 QY 1283 PTHAGAGSTGDRD-----LDRLASA-----APARAAA----- 1312
 Db 1329 -DHGOTSVSVYDMDRRFEVLEFARHRPLPDEIAGATAEARSSEGRPALAQLAALCD 1387
 QY 1313 -----LEAFLRTQVSOVL-RTPEIKVGAELFTRLGMDSLMAVELNRIRIESLKLUST 1365
 Db 1388 GREERHLAHLIRABVAAVLGHQDDAIDRDRAFRLDGFDSMTAVDLRRRLAAVTGVRREA 1447
 QY 1366 TFLSTSPNIALAQLNLALATLSLERYAAE-----NLBAGVON 1405
 Db 1448 TVFDFPHPTITRLADHYLERLVGAABAEQAPLVREVPKADDPPIAIVGMACRFPGVAN 1507
 QY 1406 -----DFVSSG-----ADDDMEIIL 1421
 Db 1508 PELMEFIVGGDAVTEPTBGMDDAL 1536

RESULT 5
 ID NCAS MYCBO
 NCAS MYCBO STANDARD; PRT: 2110 AA.
 AC 002251;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Myceroctic acid synthase.
 GN MAS.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxId=1765;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=92406887; PubMed=1527058;
 RA Mathur M., Kolatukudy P.E.;
 RT "Molecular cloning and sequencing of the gene for myceroctic acid
 RT synthase, a novel fatty acid elongating multifunctional enzyme, from
 RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guérin.",
 RL J. Biol. Chem. 267:19388-19395(1992)
 CC -1- FUNCTION: CATALYZES THE ELONGATION OF N-FATTY ACYL-COA WITH
 CC METHYLMALONYL-COA (NOT MALONYL-COA) AS THE ELONGATING AGENT TO
 CC FORM MYCOCEROSYL LIPIDS.
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE.
 CC -1- SUBUNIT: HOMODIMER WHOSE MONOMERS PROBABLY HAVE A HEAD TO TAIL
 CC ARRANGEMENT.
 CC -1- SUBCELLULAR LOCATION: Membrane-associated.
 CC -1- SIMILARITY: Contains 1 acyl carrier domain.
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M95808: AAA25369.1; -
DR PIR; B44110: B44110.
DR HSSP: P73283; 1ESM.
DR InterPro: IPR001227; Ac transferase.
DR InterPro: IPR002085; Adh_zn_family.
DR InterPro: IPR000794; ketoacyl-synt.
DR InterPro: IPR006163; PP bind.
DR InterPro: IPR006162; Panthe. attach.
DR Pfam: PF00698; Acyl_transf; 1.
DR Pfam: PF00107; ADH_zinc_N; 1.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF02801; ketoacyl-synt_C; 1.
DR Pfam: PF00550; PP-binding; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 1.
DR PROSITE: PS00066; B KETOACYL SYNTHASE; 1.
DR PROSITE: PS50075; ACP_DOMAIN; 1.
KW Fatty acid biosynthesis; Multifunctional enzyme; Phosphopantetheine;
Transferrase; Hydrolase; Oxidoreductase; Ligase; NADP; Membrane.
FT DOMAIN 1 430
FT DOMAIN 533 852
FT DOMAIN ? ?
FT DOMAIN ? ?
FT DOMAIN 2026 2096
FT ACT_SITE 177 177
FT ACT_SITE 623 623
FT NP_BIND 1561 1578
FT NP_BIND 1765 1780
FT BINDING 2059 2059
SQ SEQUENCE 2110 AA; 225577 MW; C658215D7155300 CRC64;
Query Match 25.7%; Score 1852.5; DB 1; Length 2110;
Best Local Similarity 29.5%; Pred. No. 6.Se-90;
Matches 524; Conservative 176; Mismatches 519; Indels 555; Gaps 30;

13 PIAIVGASGRLPGGVIDLSGFWTLLEGSRDTVGRVPAERWDAAMFDDPDAPGKTPVTR 72
7 PVAIVGICGRLPGGINSPPDKLWESLGRDDLVTETIPDRWDADDYDPEGVGRSVSRW 66
73 ASFLSDVACPDASFFGISPREALRMDPAHRLLEVCHEALENMAIAASALVGTETGFIG 132
67 GGFIDVDAAGFDAPFFGISSEBATSIDPOQLLETSEALEHAGDLASLAGSSTAFTG 126
133 IGPSEYEAALPQATASAEIDAHGGL-----GTPSVGAGRTSYALGLRGPVCAVDTAY 185
127 LTHEDYIVLTTHA-----GGLSPYVVTGAINSVASGRTAHTIGLGRPATFTTAC 177
186 SSSLVAVHLACQSLRSEGCSTALAGVSLMSPETLVMSKTRALADGRCKAFSAEADG 245
178 SSGIMAVHLACRSIHDERDLALAGCAVLLPFAVCVAASAQGMWLSSTGRCHSPFDADG 237
246 FGRGEGCAVVVLRKSGARADGRIILVIRGSAIINHGCASSGLTPVPGSOEIVLKRALA 305
238 FVSEGGAMVLLKRLPALRDGNRIFAVVRGTATNQGRTETLTPSEDAQVAAYRALA 297
306 DAGCAASVGVVEAHGTGTTGADPIETIOALNAVYGLGRDVATPLLSSVKNLGHPEYAS 365
298 AAGQPEPTVGVEHGTGTPIGDIPIEYRSLARYGAG-----TPALOSAKSMGHSHTASA 353
366 GITGLKVLVLSLOHGQIPAHILHAQALNPRISWGDRLVTTRATTPWDMN--TPRAAGVS 423
354 GTVGLIVAILSLRIGVVRPLLFHNRLLDELSDVETGLFVPOAVTPWPGNDHTPKRVAS 413
424 SFGSGGNNAVHVEEAATCTPRAPRPAE-----LVLGARTASALDAQAARLDLET 479
414 SFGSGGNNAVHVEEAATCTPRAPRPAE-----LVLGARTASALDAQAARLDLET 479
480 YPSQCL--GVAFSLATTRSAMERHLAVAASTR-----EGLRAADAAAGQTSFGAVRSI 533
474 H-QDCVAASDLAYTLARGRRRPPVTRAVVANLPELYEGRLREVADGALDAAG----- 527
534 ADSRGKLAFLFTGQAGQTLGMGRGLYDVVSAFREAFDLCYRLFNQSLDRPLREVMVAP 593

528 -HGDRGPV-WVPSGGSGSQAAMGTOLL-----ASEPVFAATI-----AKLEPVIAAS 573
594 A-SVDAALDQTAFT-----OPALFTFEYALAA-LWRSGVPEBELVAGHSIGELVACVA 646
574 GRSVTALTAQQTIVTIDKQVAVFVAVQVLAATMTQTYGVRRGAVGHSIGESAAVA 633
647 GVFSLDEAVFLVAARGRLMOALPAGAVNSIEAPEADV-AAAVAPHAASVSIANAAPQ 705
634 GALTLEDARVLCRSKLTMTRIAGAGMGSVLEPAQVNSELMARIGIDVVVSVASPOS 693
706 VVIAGAGQVYHIAAAMARGARTKLYSHAFHSPLMAPMLEAFPRVAVESYRRPSV 765
694 TVVIGTSDTVTRDLIAREQRDVAVREAVDVASHSPVDPIDDLALALADIAPMTPKVP 753
766 LVSNLSGKACTDEVSAPGVVVRHAREVRFADCVKLAHAAGATFEVGPSTLLGLVPA 825
754 YVSATLPDRREQVCCGAYVVDNLRKTTVQPAALVQAMMDGTRVFADVSFHLTHAVQ 813
826 C-MPDARPALIASRAGDEPATEVLEALGLVAVGVLVSMAGLFPSSGRVPLPTYMQ 883
814 TGRSLMSVAALAGMRREQPLPHGLRGLLTLLHRAAGALDSALYPA-GRVDAPLPAWT 872
884 RERYVID----- 890
873 HAALFTDDGQEQAGACTTVHPLLSGHVRLTEEPERRHWQGVTSVLMSLSDHQYN 932
891 ----- 890
933 NVAAALPGAAVCEMAALAAAEVGEALAVNDITFEQMLLDEQTPIDAVASIDAQVNVFT 992
891 -----TKADDAARGDRAPG----- 905
993 VETNRDGETTRHATALRAAEDCPRPYDITALLQHPHAVGTAMRESFAERGVTLGA 1052
906 ----- 908
1053 AFGGLTAAHTPRPREPRCPMRSCRPRSSSRAPESTRCTVLSRSARASRYGHW 1112
909 -----DEVEEGC---AVRG--- 919
1113 PAVAVGACAPACLRATVPCPRLHNAVQGLQRDPWEMADLVLDHGIVLLAVLRNG 1172
920 -----GDR-----SARLHP----- 930
1173 TGTSEDERDRLVSEBRLTLGMQORALPEVGDEAGSMLLIDTSNADVTPDMLASTLTA 1232
931 ----- 930
1233 LKSHGQGTecasLMSVODTPPNQAGLEKLSQLRGDDGVIVYGRVGDPEHSLIA 1292
931 -----PPSGRREKVEAAGDR----- 946
1293 GRQVHLVRIYTELAEFEGELPRLFVITRQAIYVPHOSGEANLEQAGRLRLAVIS 1352
947 ----- 946
1353 EHPMLRTLLIDVDEHTDVERVAQQLSGSEEDETAMRNGDWYARLTSPFLGHEERTAV 1412
947 -----PFLRLEIDPGLVHLVLRTERARPGIAGEVIAVDAAGLSFNDVQALAGVP-- 998
1413 LBDHDMGRVQVRRPDDLQLEFVASDRVPPGGQLEVAASMSINPADVLIAGFRPII 1472
999 DDLPGKPNPPLLIGGECAGRIIVAVGCVGLVVGQVIALS--AGAFATHTVTSAAVLPR 1057
1473 DD--REPQ---LGMFVGVYTAAGSVYGHQVGRVAVGSGSEGCRTPLTCANLAVLT 1526
1058 POALSAIEAAMVAVYLTAWAYLADRIARLQGERVLIHAATGGVGLAAVQMAQVGAEVH 1117
1527 PEGLTDEQATTAATATATAYGINDLAQIKAGDKVLIHSATGCVGOAASISARAKAEIF 1586
1118 ATAGTPEKRAYLESIGRYVSDRSRPFVADVAMVAGBEVDVNVNLSGSELDXSFNLL 1177
1587 ATAGNPAKRAMLRDMGVHEHYDERSVEFAEQRIRDTDYGVDIVLNSLTGAAGRALELL 1646

QY 1178 RSHGFVELGRDCYADNQLGRPLRLNISFSLVDLRGMLTERPARVALTEELGLTAA 1237
 DB 1647 PFSGFVEIGKADVYGNTRGLGFPFRGRLTFYUULALMSYVQPRVRELLATVTKLRAD 1706
 QY 1238 GVTFPPPLATLPIAVALAFAFSMAQAQHLGKVL 1271
 DB 1707 GVLTAPQCTHYFLAADAIRAMSNAEHTGKVL 1740

RESULT 6
 PSB_MYCTU STANDARD; PRT; 1538 AA.
 ID PSB_MYCTU
 AC Q10978; OS3234;
 DT 01-OCT-1996 eRel. 34. Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phenolphthalein synthesis polyketide synthase psb.
 GN PSB OR RV2932 OR MT3002 OR MTCY338.21 OR MTV011.01.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaia F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Stulson J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonyak J.F., Meyer M.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: POLYKETIDE SYNTHASE INVOLVED IN PHENOLPHTHICEINOL
 CC SYNTHESIS.
 CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETINE (By
 CC similarity).
 CC -1- SIMILARITY: Contains 1 acyl carrier domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AL021070; CAA15929.1; -
 DR EMBL; AE007122; AAK47329.1; -
 DR PIR; E70874; E70874.
 DR TIGR; MT3002; -
 DR TubercuLac; RV2932; -
 DR InterPro; IPR001227; AC transferase.
 DR InterPro; IPR000794; Ketosacyl-synt.
 DR InterPro; IPR006163; PP_bind.
 DR InterPro; IPR006162; Pantne_attach.
 DR Pfam; PF00698; Acyl_transf; 1.

DR Pfam; PF00109; ketoacyl-synt; 1.
 DR Pfam; PF02801; ketoacyl-synt C; 1.
 DR Pfam; PF00550; pp-binding; 1.
 DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; FALSE_NEG.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 DR PROSITE; PS50075; ACP_DOMAIN; 1.
 KM Hypothetical protein; Multifunctional enzyme; Oxidoreductase;
 KM Transferase; NADP; Phosphopantetheine; Complete proteome.
 FT DOMAIN 145. 1495
 FT BINDING 1458 1458 ACP_CARRIER (ACP).
 FT SEQUENCE 1538 AA; 162527 MW; B55E2A2042AD00CC CRC64;
 SQ

Query Match 22.1%; Score 1592.5; DB 1; Length 1538;
 Best Local Similarity 31.2%; Pred. No. 2.2e-76;
 Matches 488; Conservative 207; Mismatches 600; Indels 271; Gaps 44;

QY 3 DRPIERAEDPIAIVGASCRPGYIDISGFWTLLEGSRDVTGVPAERWDAAMFDDP 62
 DB 25 DVSRIAAVEPVAVGVGICRFPDGDGPEFMDFLVGRNASTVPAERWDAEAFYHDDP 84
 QY 63 DAPKTPVTRASFIDVACPDASFPGISPRRLAMDPAHRLLLRVCWELNMAIAPAL 122
 DB 85 LTPGNTTKGCGFVDVAGPDAEFEGITPRDAAMDPOKMLVEALSHAGIPPSL 144
 QY 123 VGTETGVFIGIPSEYEALPOATASAEIDAAGLGTPSVGAGRISYALGRPCVAVD 182
 DB 145 GGTETVAVMGVYFNEYQSM--AASPQVDAVSGTGNHSTVGRISTYLLGRPAVAVD 202
 QY 183 TAYSSLYAVHLAQSIRSGECSTALLAGVSLMSPSTLVLSKTRALARDGCASAE 242
 DB 203 TACSSLYAVHLAQSIRSGECSTALLAGVSLMSPSTLVLSKTRALARDGCASAE 262
 QY 243 ADGGRGRCGAVVVKRLSGRADPDILNIRSATINHDASGLTPNCSOETVLR 302
 DB 263 ADGVRGEGAVVVKRLTDVNRDQVLAVRRSAAVQDSRSNGVAPNPAQCDVAVD 322
 QY 303 ALADAGCAASSGVGEAAGTGTGDPLEIGALAAVVGCLGRDVAFTPLIGSKTNLGHPE 362
 DB 323 ALRSGDVADESVNVVEAHGTGTVDGPIEPFALAAATGHGDACA--LGVKNIGHLE 379
 QY 363 YASGITGLKVLVSLQHQIPAILHAQALNFRISGDLRLTVTARTFPMWMTPRRGV 422
 DB 380 AAGIAGIKATLAVOPARTIPNLFPSQWPAIDAASTFPPQNSWEPFAEGRRAAV 439
 QY 423 SSFMSGTHNVAVLE-APATCPPARPARAEILLVSARASLLDQAARLRLHLEYR 481
 DB 440 SSFELGTHNAVILIEQSESLAPVSEGEDGVSTLVVTKTAQRAAATAQVLAAMMEGPG 499
 QY 482 SQ-CLGDAVAFSLATTRAMEHRLAVAAATSRGLRAALDAAGOTSFGAVRSIADSRGK 540
 DB 500 AEVAVADVAVHNVNHRARQATFGTVVARDRAQALAGRLAAGHAGGV-SHODGSFGR 558
 QY 541 -LAFLTGQGAQTIGMGRGLYDVMSAFREAFDLCVRLFENELDRPLREVMMAEPASVDA 599
 DB 559 GTVFVYSGRSGQMGMRQLADAPAAVALEPVEVQAGSLRDVI--ATGKELV 615
 QY 600 LLDQTAQTPRALFFFEVALAALWRSKVEPRLVAGSHIGELVAACVAGVSEDAVFLVA 659
 DB 616 GIEQ--IQLGLGMQTLLELMRSYGVQPLVIGSHGVEVAALVAVGALLPAGGLRATA 672
 QY 660 ARGILMOALPAGAMVSEIAPADVAVAVPAHSAVSIAAVNAPQVVIAGAGPVAHA 719
 DB 673 TRAILMAPLSGGGSMALLGDAATELLADY-PQVTVGIRNSRQVVIAGPTQIDELI 731
 QY 720 AAMARAGARTKALHVAHSHAFHSPLMAPMLEAFGRVAESVYRPSIVLVSNLSGACTDEV 779
 DB 732 ARVAAQRFAFSRVNIEVAPHPAMDALQPMRSELDLTPRTPIGIIISTVYADLHPPI 791
 QY 760 SSPGVYVRAHREVVPADGVKALHAAGAG-----TFVVGKSTLLGLVAPACMPDA-RP 832
 DB 792 FDAEHMAITNNKPNRFO--QALASAGSGADGAVHTTEISAPDLTQALADTLEDAHRP 848
 QY 833 A-----LASSRAGRDEPATVLEALGILMAVGLVAGLFPSCGRRPV----- 876

QY 118 ABSALVGTETGVFIGIGPESEYEALPQATAS--AEIDAHGLGTMPGVGAGRISYALDGLG 176
 DB 208 PGTLRRSATGVFAGACSEYGA---MADADLSQVDMGMSNGAMSIIINRLSYFLDLNG 264
 QY 177 PCVAVDTAVSSSLVAVHLACQSIIRSGECSTALAGVSLMSPELTWLSKTRALADGRC 236
 DB 265 PSVAVDTACSSSLVAVHLACQSIIRSGECSTALAGVSLMSPELTWLSKTRALADGRC 324
 QY 237 KAFSAEADGFGGEGCAVVVLKRLSGARADGRIILAVITGSAIINHGSAGLTVPGSSO 296
 DB 325 RAFDADADGVRGEGGVVVLKRLTDQRDGDVLAIVIGSAVNOGSRNGMLAEPAPAO 384
 QY 297 EIVLKALADACCAASSVGVVEAHTGTTLGDPTEICQALNAVYGLGRDVAATPLIGSVKT 356
 DB 385 MAVLRAAYTNAGQPEVDYVEAHGTGTLGDPTEIARALGTVLGRPREDSPLIGSVKT 444
 QY 357 NUGPEYASGITGLKLVLSLQHQI PAHLHAQALNPRI SWGDLRLTVTRAPTWDNMT 416
 DB 445 NIGHEAAGIAGFIKTIVLAVOHQI PRNQHETAPPHI PFTDLRMKVVDQTQEWPAICH 504
 QY 417 PRAGVSSFGMGSTNAHVLEEARCTPRAPERB--ELVLSATASALDAQARLAD 475
 DB 505 PRAGVSSFGMGSTNAHVLEEARCTPRAPERB--ELVLSATASALDAQARLAD 564
 QY 476 HLETPSO--CLGDVAFSLATTRSAMEHRLAVATSREGLRAALDAAGQOTSPGAVRSIA 534
 DB 555 WMEGCAVDVALADVANTLNHHSROPKFCTVVARADDTQAIAGRALAAGQHPGVNP-A 623
 QY 535 DSSRGK-LAFLETTGCAQTLGWSRGVLYDWSAFREAFDL CVRLFNGELDRPLREVMABP 593
 DB 624 DSSPGCTGVFVYSGRSQWAGMGRQLADEPAFAAVALLEPVFVQAGFSLHDVL---- 679
 QY 594 ASVDAALDQNTFTQALFTFEYALALMRSMGVBEIVAGHSIGTLVACVAGVSLSD 653
 DB 680 --ANGBELGIDQIQLIGMQLALTELCYGVPRDLVIGSMGVAAVAAGALTPE 737
 QY 654 AVFLVAARGLMQALPAGG--AMVSIAPADVAAVAAPAAVSIAVAAPQVYIAGA 711
 DB 738 GLRVATRRLMAPLISGQGMALDELADPTTEALINDP---QVTLGITSNRQVYIAGP 794
 QY 712 GQPVHAIAAAMARGARTALHVSHPPLMAPMLAEFARVAESVSRPSIVLSNLS 771
 DB 795 TEQIDELIARVRAQRKFAVRVNI EVAPHPNADALOPMRSELDLTPRTIGISTTY 854
 QY 772 GKACTDEVSFGVWVHAREVVREADGVALHAAGG-----TFEVRP-----KSTLLG 821
 DB 855 ADLHTOPVDAEHMATNMENPVRFQ---QALASAGSGADGAYHTFEISAHPLLTQALID 911
 QY 822 LVPACVDPARPALLASRRAGDEPATVLEALGGLMVGGLVSWAGLP-----SGGRVP 876
 DB 912 TLHSAQPGARYSLSGLQRTDDVYFRNLNKAKHTI-----HPHTHPRPPHP 962
 QY 877 LFTYVQEREWIITDYADDAARGDRRAPAG-----HDEVEEGAVRGDRRSARL--DH 929
 DB 963 IFTTWMQHRHMITTYKYPAGSVGS--APRAGTLGQHTTVATVSAPRSHLMQARLAPDA 1020
 QY 930 PPEEGRR-----EKVEAA-----GDRPRLIDEPRV 957
 DB 1021 KYQGGHREHOVEVPAVSVLHTLITSAATELGYSALSEVRFPQPIFADPRRLIYV---V 1076
 QY 958 LDHLVLVTERRA-----P 971
 DB 1077 ADNRAISLASSPRACTPRDRWTRHVTACQSSSPSDASSLNHHRANQPRERARHDLIP 1136
 QY 972 GLGEVEIADAAGLSF-----NDVQALAGVPPDLPOKPNPRLILGECBAGRIY 1020
 DB 1137 DIAELLAMRGIDGLPESMTVASWTQSSNLVAILD--PEALPGSGTGPL--DAVHIA 1192
 QY 1021 AVGEGVNG--LVVGQPIALASAGAFATHTVTSAL-----VLPRPALSAIEAA 1068
 DB 1193 ALSDVADSRLVYPASTIEQISLQGVVTVGRSSVTLNTFAHDDGITYDVTVAAHGEVPSLS 1252

QY 1069 M-PVAVYLTAWVALDRIRALQGERVLVIAHATGVC-----LAAYON-----AQRV 1112
 DB 1253 KSLIRYALDFGLD--VGAQP-----PASTPVEAYCOATNFVHTIDMQPTVPDATHP 1305
 QY 1113 GAE--VH-----ATGTPKAYLESIGVRYVSDRSRDFVD----- 1148
 DB 1306 GAEQVTHPEPAVIAIDDDGALLCETLEGAGYOPAVMSDQARVYVYVADSDPAGADETD 1365
 QY 1149 ----VRAMTGESE----- 1157
 DB 1366 VDPFAVICTEITGLVTRTAEADADKPAALMILTRGVHESVADSALRQSLKLAGYIAE 1425
 QY 1158 -----VDVNLISGELIDKSNLRSRGRFVELGRD----- 1190
 DB 1426 HEMGLGLVDLAINDDLEFPGFALBELAKPSKSI--LVARDQVLA PALAPVGEPAKRS 1484
 QY 1191 --CYAD-----NOGL-----RPLRNLFSVLVD----- 1212
 DB 1485 IQCRDPAVYLTGGLGALGLMADWLADRGARHLVLTGRTPLP PPRRDWQDLDTLDELRRR 1544
 QY 1213 -----LRGMLERRPAR-----VRAL-----EELGLI----- 1235
 DB 1545 IDAITALEKRGVTEAVALADNGCREDOQLALARDGCAPIRGIIHAAGITNDOLVISM 1604
 QY 1236 -----AAGVFTPPP-----IATLP 1249
 DB 1605 TGDVAQVWMPKIGSOVLHDAFPFGSVDFYLTASAGIFCI PGGGSYAAANSYLDALA 1664
 QY 1250 IARVADAFRSM-----AOQHLKLVLTGLDPE----- 1277
 DB 1665 RARROOGCHTMSLDVNAWRGLGLADAOI VSEELARMGSRDITPSEAFPAWEFVDGYDVA 1724
 QY 1278 --VOIRIPTHAGAPSTGDRDLIDLASAPAR--AAALEAFRTQVSQVLRTEIKVG 1333
 DB 1725 QAVVPMAPACADSGANAYILPARNWSVMATVRSLEGLRIITIAELRVEKEKD 1784
 QY 1334 AEALFTRLGMSLMAVELRNRIEASLKLSTTFLSTSPNIALAONLIDALATLSLER 1393
 DB 1785 TDRPFAELGNSLMAVMAIRREAEQFVGIELSTMLFNPHTVASLASYL-----AKR 1835
 QY 1394 VAAENLRAGVQNDPVSQA 1412
 DB 1836 VAPHDVSDNDQISALSSSA 1854
 RESULT 8
 WA EMENT
 ID WA EMENT STANDARD; PRT; 1986 AA.
 AC 003149;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Confidial Green pigment synthase (EC 2.3.1.-).
 GN WA.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 CC Eurotiales; Trichocomaceae; Emericella.
 OX NCBI_TaxId=162425;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93101122; PubMed=1465094;
 RA Mayorga M.E., Timberlake W.E.;
 RT "The developmentally regulated Aspergillus nidulans wa gene encodes a
 RT polypeptide homologous to polyketide and fatty acid synthases";
 RL Mol. Gen. Genet. 235:205-212(1992).
 CC -FUNCTION: THIS PROTEIN CONDENSES CARBON UNITS TO FORM AN
 CC INTERMEDIATE YELLOW POLYKETIDE PIGMENT THAT IS POLYMERIZED
 CC BY CONFIDIAL LACCASE TO FORM THE GREEN PIGMENT IN MATURE
 CC ASEXUAL SPORES (CONIDIA).
 CC -CORCTOR: CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTHETINES
 CC (POTENTIAL).
 CC -PATHWAY: Confidial green pigment biosynthesis.
 CC -SIMILARITY: Contains 2 acyl carrier domains.

	Score	DB 1	Length
Query Match	16.4%		1986
Best Local Similarity	26.9%	Pred. No. 1.6e-54	
Matches 419; Conservative 242; Mismatches 635;			
			Indels 264; Gaps 53;

```

OY      8  RAEDPIIAVASCLEPGVIDLSEFWTLLEBSRTVRVAERDDAAMFDPDDADGK 67
Db      3 13 RAELSKIAIIGMSGFRP- EADSPODFMULLKGLDVHRKVEBDRMDADAHYDLTGTAINT 431
OY      68  TPVTRASFSDVACEFASFFGISPREALRMDPAHRLLEVCWEAEALENNAIPASALVGE- 126
Db      4 32 SKVPYGCWIRERGLDFDRPFNMSPREALQADPAQGLATALLTVAYEALLEGACFVPDSTPSTOR 491
OY      1 27 --TGVFITIGSEYEYALPQATASAEIDAHGGLGTMPSVSGRISYALGLGPCYAVUTA 184
Db      4 92 DRVGIFYEYMTSDDYR---EYVNSGDDITUYEYFGGNRAFPTRIRINUYEFKFGSPVSVDTA 547
OY      1 85 YSSSLVAVHLACQSLRSECESTALAGVSLMSPSTLWLSKTRALARDCKCAFSABED 244
Db      5 48 CSSSLAAIHLACNSTWRNDCDTAIGVGNILTPNHNHAGIDRGHLSHTGNCNTEFDDCAD 607
OY      2 45 GFGRGEGCAVYVVKRLSGRADGDRILAVIRGSAIINHDCASSGLTVPNGSOSYVLRAL 304
Db      6 08 GYCRADGEGTVVVKLELADLANDPBILGVINGCAYTNHSAEAVSYIRPYPVGAQAFIKKLL 667
OY      3 05 ADAGCAASVGVAEHAGTTLGDPIELQALNAVYGLG--BDVATPLLIGSKTNMLGHPRE 362
Db      6 68 NEANVDPKNIISTYLEHGTQOAGDAVENQOSVDVAPDRHRKPGOSLHGSKSNITGICE 727
OY      3 63 YASGITGLLKVYLSLOHGOIPAHLYAOA-LNDRISWGDRLTVTARTPRPMDWNP 417
Db      7 28 SASGVTSLVKYLMLMKEMNIPRHGIKTKIKHNHPTDLAQNRVHIALQ-PAWNRPSFGK 786
OY      4 18 RRAGVYSFGMSGSTNAHVULEAPAACTCPRAPE---RPAELIUSATTAAL---DAQ 469
Db      7 87 ROIPLNTPNSAGGNATLLLEDGPVSD-----PEGEDKRTYVITLSASOTRLONNIDL 841

```

QY	470	AARLRLHLEYT---	PSQCLDPAVFAISLATTS	AMEHRLA	VAATREG	IRALDIAAQC-	524																																																	
Db	842	QYI7SEBEXTFGV	KD5NALPS	LAVTTARK	IINHPRTAL	AGSSFGQERD	LIASSREFEV 901																																																	
QY	525	---TSBGA	VR5IADSSRG	KLAFLFTG	QAGQTLGMR	GLDVMSAFRE	FDL--CYRLFN 578																																																	
Db	902	AVAPAKT	PG-----	IGFLFTG	QAGQVAA	MGKQLYEDCS	HFSALEHNDICISQ -G 948																																																	
QY	579	QELDRPL	REVMKME	RPASVDA	LLDQTLF	QPALFTFE	YALALMRSGVRE	PELVAGSISG 638																																																
Db	949	QDLPS	TLPLVDG	SLPSLETS	PPVVQQLTT-----	CYVMA	LSSTFPA	SIGITPSPFLG	ISLG 1003																																															
QY	639	ELVA	CACVAGF	SELEDA	FAFLVA	ARGRLM--CAL	PRAGAMVS	IEAREAD	VAAVA	PHASVS 636																																														
Db	1004	DFAMA	NAAGVL	STSDT	LYACGR	RAQLTER	COGTHA	MLIKAP	LV	EVKQLT--NEKVHD 1061																																														
QY	697	IAA	NAVPDO	VVIAC	AGOPV	HAIAA	MAAR	KARTKAL	HSVASH	AFPLMAP	LEAFGR	VAES 756																																												
Db	1062	MAC	INSPE	ETVIG	SPKSS	IDE	LSRACE	KEQIKST	ILTPVAF	HA	QVEP	FLIEBLE	LEKALOG 1121																																											
QY	757	VS	RRPS	ITIV	SNLS	NGACTD--E	VSS	PGVWR	HA	AEVVR	FP	ADGV	ALHA	AGAG---T 809																																										
Db	1122	IT	FKKSV	VPV	SALL	GEVITE	AGSN	IINA	EYLVR	HC	ETYN	FL	SA	FAVA	NALG	GGQTL 1181																																								
QY	810	FVE	EPK	ST	LLGL	VAP	CM	PAR	PALLAS	BS	AGRE	PA	TVLE	ALG	GLWA	VG	SWAGL-- 867																																							
Db	1182	WLE	VG	PH	TV	SG	MKAT	L--G	QOT	TMA	SLR	DE	D	TW	KVLS	NS	LS	TY	L	AG	V	DN	IM	KOYHQ 1240																																
QY	868	-F	PS	GGR	R	V	LP	TY	PMQ	RERY	TD-----	TY	AD	DA	R	DR	AP	RG	CH	NE	VE	EG	GV	VRG 919																																
Db	1241	D	FS	SH	H	V	L	P	LP	TY	KW	L	KW	M	Y	P	Y	R	N	F	C	L	T	G	S	S	S-----	A	S	A	S	L	O	P	T	F	L	T	S	KO	V 1296															
QY	920	G	D	R	S	A	R	L-----	D	H	P	P	E	S	E	G	R-----	R	E	K	V	E	A	G	D	R	P	R	L	E	I	D	E	G	V	D	H	L	R	V	T	E	R 968													
Db	1297	V	E	S	D	D	G	L	T	V	V	A	N	D	I	A	P	D	L	R	V	I	Q	H	K	A	G	A	L	C	P	S	L	V	A	D	S	A	Q	T	L	A	E	V	L	---EK 1353										
QY	969	R	A	P	G	L	G	E	V	E	I	A	N	D	A	G	S	F	N	D	V	Q	L	A	G	M	P	D	D	L	P	G	K	N	P	E	L	L	G	E	C	A	G	I	V	A	G	E	V	Y	G 1028					
Db	1354	Y	K	P	E	L-----	K	O	S	G	L	D	V	C	N	V	TV-----	P	K	P	I	A	K	T	G	K	O	F	A	I	S	A	T	A	N	V	W	D	K 1394																	
QY	1029	L	V	E	G	O	P	I	A	L	S	A	G-----	A	F	T	H	V	T	S	A	A	L	V	L	P	R	O	P	L	S	I	A	E	A	N	A	M	P-----	1070																
Db	1395	H	V	S	O	V	S	V	T	R	A	E	G	K	L	I	D	H	A	C	E	V	K	L	P	D	C	M	A	A	D	L	E	M	G	R	G	S	V	L	V	K--R	E	I	E	L	L	E	N	S	A	V	A	G	D	A 1453
QY	1071	---VA	L	T	W	A	Y	A	L	D	R	I	A	L	O	G	E	R	V	L	---H	A	A	T	G	V	G	L	A	A	V	O-----	W	A	O--	1110																				
Db	1454	H	R	L	R	G	M	Y	T	K	F	A	L	V	D	E	N	Y	O	G	I	R	E	V	I	L	D	S	E	H	E	A	T	A	L	V	T	K	O	A	P	A	N	F	H	R	N	P	Y	I	D	S	F 1513			
QY	1111	-H	V--GA	E	V	H	A	T	A	G	P	E	R	K	A	T	I	E	S	I	G	V	Y	V	E	D	S	R--D	R	F	A	D	V-----	R	A	W	T 1153																			
Db	1																																																							


```

QY 1008 PLIL-----GGECAIRIVAVG----- 1023
DB 1137 ALMPMPQIERVEVFTGDPDPKIMLVQGEASDSVPISHSVSEAGEVLAQKFMNPFSEI 1196
QY 1024 EGVNGL-----VVGQPVIALSAGAFTHVTTAALVLPPOALS-AIEAAMP----- 1070
DB 1197 EGTGPGVSGMESLVHQ--IAMPATPAEPELSIEIVLVSPDATTALVAPASHPTRVNSP 1254
QY 1071-----VAYL-----AMVLDRIARLPGE-- 1090
DB 1255 QESSTOEFPSSNASLPLEKTVVTTYIPEVASIAEVPASESTWMLLEIKTVANGSLP 1314
QY 1091 -RVLIHAATGCVGLAIVQMAQ-----HYGAEVHATAGT-----PEKRAYLESL 1132
DB 1315 IKVFTLTANIGEGOPTALAQSPLYGIAVIASEHPDGLTIVBEVPVPLSTMRYIQGA 1374
QY 1133 GVRVSD--SRSDRPVADV-----AMTQSGCV----- 1158
DB 1375 DIRINDGIARTSRFSRLPKNLLPASEGRLPREGTYLITGGGLGLGEVADFLVEK 1434
QY 1159-----DVVNSLSGELIDKSFNLSHGRFV-----ELGKRCQYAD 1194
DB 1435 GARRLLISRRLPRTTWQVSEDLQPTI--AKIRLESRGASVHVLPIDITKPDVVEQ 1492
QY 1195 NQGLRPFLRLSF-----SLVDRG-----NMLESPARVALLLEELGLIAGVFTPP 1243
DB 1493 ----LTTALDRLLSPVQGVVHAAGVLDNELVWQTRDAFNRIAPRIAGALALHVEFPP 1548
QY 1244 P-----IATLPI--ARVADAF----- 1257
DB 1549 KSVDFPVMSSCGNLVFTGQASVSGNAFLDTLATRHARLGDAAVSPQWTSWRGLMGGA 1608
QY 1258-----RSMAGQOHLGKL-----VLTGDPPE 1277
DB 1609 STDFINAELESKGITVTDTRDEAPRAWOHLAKYMDHCIVLRSAFEDGEPIPVISILNDIA 1668
QY 1278 VQIRIPTHAGAPSTGRDRLDLRLASAPARAALAEPLRTQVSOVLRTPEIKVGEAL 1337
DB 1669 VR-RVGVSVNTSPAAAGSS--DAVPTSGPELK-AVDEKIRGCYAKVLCQMTADVDVSKAA 1724
QY 1338 FTRLGDSLMAVELRNRIEASLKUKUSTTSLTSPNIALIAQNLIDALA 1386
DB 1725 LADLGVDSTVTLRQLQTLTKIAVPPTLVWSPHTVSLHVAWFAEKLA 1773

RESULT 10
PKSI_ASPPA STANDARD; PRT; 2109 AA.
AC Q12053;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aflatoxin biosynthesis polypeptide synthase (PKS).
GN PKSL1.
OS Aspergillus parasiticus.
OC Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitospotic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5067;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=NRRL 2999;
RX MEDLINE=96042102; Pubmed=7592391;
RA Peng G.H., Leonard T.J.;
RT "Characterization of the polypeptide synthase gene (pksl1) required
for aflatoxin biosynthesis in Aspergillus parasiticus.";
RL J. Bacteriol. 177:6246-6254 (1995).
CC -1- FUNCTION: INVOLVED IN THE SYNTHESIS OF THE POLYPEPTIDE NUCLEUS OF
AFIATOXIN FROM HEXANOYL COA AND SEVEN MALONATES.
CC -1- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTHETHEINE (By
similarity).
CC -1- PATHWAY: Aflatoxin biosynthesis, first step.
CC -1- SIMILARITY: Contains 1 acyl carrier domain.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
DR EMBL: LA2766; AAC1675.1; -
DR EMBL: LA2765; AAC1674.1; -
DR PIR: T17490; T17490.
DR InterPro: IPR001227; Ac. transferase.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR006163; Pp. bind.
DR InterPro: IPR006162; Pplant. attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00698; Acyl_transf. 1.
DR Pfam: PF00109; ketoacyl-synt. 1.
DR Pfam: PF02801; ketoacyl-synt. C; 1.
DR Pfam: PF00550; Pp-binding. 1.
DR PROSITE: PS00975; Thioesterase; 1.
DR PROSITE: PS50075; ACP DOMAIN; 1.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; FALSE_NEG.
DR TRANSFERASE; Acyltransferase; Phosphopantetheine;
KW Multifunctional enzyme.
FT DOMAIN 374 805 BETA-KETOACYL SYNTHASE.
FT DOMAIN 1714 1745 ACYL/MALONYL TRANSFERASES.
FT DOMAIN 543 543 ACYL CARRIER (ACP).
FT ACT_SITE 543 543 THIOESTERASE.
FT ACT_SITE 993 993 BETA-KETOACYL SYNTHASE (BY SIMILARITY).
FT ACT_SITE 993 993 ACYL/MALONYL TRANSFERASES (BY
SIMILARITY).
FT BINDING 1746 1746 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT ACT_SITE 1937 1937 THIOESTERASE (BY SIMILARITY).
SQ SEQUENCE 2109 AA; 230715 MM; CB701372A16B8551 CRC64;

Query Match 16.3%; Score 1173.5; DB 1; Length 2109;
Best Local Similarity 27.2%; Pred. No. 3.8e-54;
Matches 409; Conservative 222; Mismatches 625; Indels 247; Gaps 54;

QY 9 AADPPIAIVGASGRLLPGVYIDLSGFVTLLEGSDIYGRPAEHWDAAMFDDPDPAKGT 68
DB 369 AGCKRLAIYVMSGRFPESPTTES--FWDLLYKGLDVCKEVRRRWDINTHVDPGKARNG 427
QY 69 PVTBASFLSDVACFDSPFGISPREALRMDPAHRLLEVCMELENAALAPSALVGTG-- 126
DB 428 ATKWGCWLPDSGDFRFRFGISPKENAPQMDPAQMALMTYEMEAAGLVPTTSTQND 467
QY 127 -TGVFTIGPSEYEALPQATASAEIDAH---GGLGTPSVGAGRISYALGRFCVAV 181
DB 488 RIGVFHGVISNDW---METNPAQNIIDTYFITGNNGFIP---GRINFCFEFAGPSYTN 539
QY 182 DTVYSSSLVAVHLACQSLASGECSTLLAGVSLMSPTLVMLSKRALAPDRCAFA 241
DB 540 DTVACSSSLAHLACQSLKWRGDDTVAVAGTMMIYPPDDHTGIDGKFFLSRTGNCPRYD 599
QY 242 EADGFRGSGCAVVLKRTSGARADGRILAVRGSAINHDGASGLTVPNGSSQGEIVLK 301
DB 600 KADGICRAEGVGVFVKRLLEDALADNDPILGVILDAKTHNSAMSEMTPHVQAQIDNMT 659
QY 302 RALADAGCAASVGVYEAHGTGTTGDPLEIQALNAVY--LGRVATPPLILGYSYKTNL 358
DB 660 AALNTTGLHPNDPSYIEHMGITGOVDAVEMESLVFAPSETPARADPLFVGSAAKAVY 719
QY 359 GHPVYASGITGLKVLVLSHQOIPHL--HAQALNPI--SWGDIRLVTVTRARTWPPW 414
DB 720 GHEGEGVGSYTLKIVLMAQHDPIPHCGIKPKSKINRFPDIDGARNVHIAFEKPKWPT 779
QY 415 NTPRRASVSPGSGTNAVHLEAPAACTPPAPERPELVILVSARTASALDAQARLR 474
DB 780 HTRRVYLNNFSAAGNTALIVEDAPERHMPTEKDRSSHIVALSAGASMTINERLH 839
QY 475 DHLETYPSOCLADGVAFSLATTSAMEHRLAVATATRSREGRLAALDAQAQ--GQTSFGAVR 531

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:33 ; Search time 49.4261 Seconds
(without alignments)
2764.850 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210

Sequence: 1 VADRIERAEDPIAIVGAS.....GVQNDPVSQADQWEIHAL 1421

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2331	32.2	4151	2 G70944	probable polyketid
2	2298.5	31.9	2103	2 G86925	probable polyketid
3	2211.5	30.7	2478	2 AH2140	polyketide synthase
4	2207	30.6	1502	2 H70984	probable polyketid
5	2136	29.6	2518	2 A12140	polyketide synthase
6	2120	29.4	1587	2 AB2012	hypothetical prote
7	2093.5	29.0	3573	2 S23070	erythronolide synt
8	2063.5	28.6	3413	2 T17467	rifamycin polyketi
9	2052.5	28.5	3739	2 T17410	polyketide synthase
10	2051.5	28.5	4613	2 T17409	polyketide synthase
11	2025	28.1	2126	2 H70621	probable polyketid
12	2014.5	27.9	2188	2 A70984	probable polyketid
13	2013.5	27.9	6260	2 T30228	polyketide synthase
14	2002.5	27.8	4735	2 T17463	rifamycin polyketi
15	2000.5	27.7	10223	2 T30225	polyketide synthase
16	1983.5	27.5	2201	2 S73014	polyketide synthase
17	1976	27.4	8563	2 T30226	polyketide synthase
18	1965	27.3	6420	2 T30283	polyketide synthase
19	1955	27.1	2100	2 T03223	probable polyketid
20	1925.5	26.7	3491	2 T43231	probable 6-deoxyer
21	1908.5	26.5	2108	2 H70819	probable polyketid
22	1900.5	26.4	2124	2 T28658	polyketide synthase
23	1889	26.2	3519	2 S43048	polyketide synthase
24	1875	26.0	3172	2 S22012	erythronolide synt
25	1869	25.9	2126	2 E70522	probable polyketid
26	1855	25.7	2111	2 A70668	mycocerosate synth
27	1852.5	25.7	2110	2 B44110	mycocerosate synth
28	1831.5	25.4	5069	2 T17464	rifamycin polyketi
29	1827	25.3	2116	2 C86926	probable mycoceros

30	1812	25.1	1774	2 T17421	polyketide synthase
31	1809	25.1	7576	2 T17428	FK506 polyketide s
32	1800	25.0	3178	2 S13595	6-deoxyerythronoli
33	1798	24.9	1562	2 T17411	polyketide synthase
34	1790.5	24.8	1402	2 D70634	probable polyketid
35	1785.5	24.8	1763	2 T17465	rifamycin polyketi
36	1784	24.7	1570	2 AC2012	hypothetical prote
37	1776.5	24.6	1762	2 T03222	probable polyketid
38	1773	24.6	1728	2 T17466	rifamycin polyketi
39	1751	24.3	1937	2 T03224	probable polyketid
40	1719.5	23.8	1017	2 B70985	probable polyketid
41	1712.5	23.8	2118	2 S72705	mycocerosate synth
42	1684.5	23.4	1293	2 T30871	orebellinic acid sy
43	1678	23.3	2297	2 T34918	polyketide synthase
44	1643.5	22.8	1827	2 B70984	probable polyketid
45	1638	22.7	1346	2 T17412	polyketide synthase

ALIGNMENTS

RESULT 1

G70944
probable polyketide synthase RV2048c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: G70944

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentile, S.; Hamlin, N.; Holtrope, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G. A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome A: Reference number: A70500; MUID:98295987; PMID:9634230

A: Accession: G70944

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-4151 <COL>

A: Cross-references: GB:AL021899; GB:AL123456; NID:93242282; PIDN:CA17262.1; PID:G789678

A: Experimental source: strain H37RV

A: Gene: pks12

C: Genes: pks12

C: Superfamily: Mycobacterium tuberculosis probable polyketide synthase RV2048c; 3-oxoacyl

homology; [acyl-carrier-protein] S-malonyltransferase homology

C: Keywords: carrier protein

F: 556-455/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>

F: 559-841/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>

F: 1680-1661/Domain: short-chain alcohol dehydrogenase homology <SAD1>

F: 1963-2038/Domain: acyl carrier protein homology <ACPI>

F: 2078-2478/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F: 2587-2860/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>

F: 3710-3891/Domain: short-chain alcohol dehydrogenase homology <SAD2>

F: 3995-4066/Domain: acyl carrier protein homology <ACP2>

Query Match
Best local similarity 32.2%; Score 2321; DB 2; Length 4151;
Pred. No. 6, 7e-117;
Matches 613; Conservative 203; Mismatches 535; Indels 498; Gaps 31;

QY	6	IRAAEDPIAIVGASCRPGVYIDSGFTLLIESGRDVGKPAER-WDAAMFPDDPDA	64
DB	2051	YRTTSEDDPIAIVGACRYFGVNSPDDMDMLIQGRDVISEPPADRGMDLQVNDPDA	2110
QY	65	PKRTVTASPLSDVACPDASFFGISPREBAHMDPAHLLFVCEBALENAIAPSAVUG	124
DB	2111	ACACTTGGFVDGPDPAFVGVPSEBALMDQHRLLELSWEALERAIDPGLRG	2170
QY	125	TETGVFIGISEVEYEAALPOATASAEIDAAGGLGTPMSVAGRIYALGLRCPVAVDTA	184
DB	2171	SATGVFAGVMVQGY-----GMPAAPVGVFRITGLSSVASARVAVYVGLSPAVSVDTA	2225
QY	185	VSSSLVAVHACQSLRSGECSTALAGVSLMISPTVLWLSKTRLLARDGRKASAEAD	244
DB	2226	CSSSLVAVHMAVGVSLRSGCDLALAGVTVNATPDIFVEFSRMRGLSPDGRCKAFAMAAD	2285

QY 245 GFGREGCAVYVLRSLSGARADGRILAVIRGSAINHDSGSLTVPNSSOEIVLKRAL 304
 DB 2286 GTFPSBGGMLVLRSLSDARRLGHPIVLAIVVGSANVDGASNGITLAPNGSQQCVVAAL 2345
 QY 305 ADAGCAASVGVYEAHGTGTLDPDPIEIQALNAVYGLGR-DVATPLLIGSVKTNLGHPEY 363
 DB 2346 ANNGLSAEVDVVEGHTGTTLGDPIEAQALLLATVGGDRGEPGPELIGSVKSMGHTQA 2405
 QY 364 ASGITGLKLVLSLQHQITPAHLHAQALNIRISMGDLRLTVTRARTWPMNTPRRAGVS 423
 DB 2406 AAGVAGYIKVYLMRHELPATLHVDPSPHVDMSAGAVELTLAPRWMPGARTRRAGVS 2465
 QY 424 SFGMSGNNAVVEEAPATCTPAPERPAELVLSARTASALDAQAARLDHLETPSQ 483
 DB 2466 SFGISGTHNAVITIEADVVRREAGMAGPVVPMVSAKSSSALRGQALAAIAYRGDGL 2525
 QY 484 CLGDVAFSLATTRSAMEHRLAVALATSREGRLAALDAAGQTSFGAVRSIADSSRGLAF 543
 DB 2526 DVADVGSILA-GRSVFEHRAVAVVGGDRDLAAGLDELADQLGSSVVRGTA-TRAGKTVE 2583
 QY 544 LFTGCGAQTLGMRGLYDVMSAREAFDLCYRLFNQGLDPLLEWMAEASVDALLDQ 603
 DB 2584 VFGQGSQWLGKMGGLHAGYVFAEAENTVVGELDRHLRLPREVMGH----DENILNS 2639
 QY 604 TATFOPALFTFEYALALMRSMGWPELVAGHSIGELVAAVAVGFSLEDVAFVLAARGR 663
 DB 2640 TEPAQPLFAVEVALFRLGSKGVPRPFVNGHSIGELSAHVAVGSLERNAVLVAARGR 2699
 QY 664 LMOALPAGAMVSIIEADVAANAAPHAAVSIAVANADOVVIAGACOPVHAIAAMA 723
 DB 2700 LMOALPAGAMVAVQAEEVEVRPL--SAEVDIAANVPASLYISGAQNAVAVAVDQLR 2756
 QY 724 ARGARFKALVSHAFHSPLMAEAFGRVASEVSTRRESIVLVSNISGACCTDEVSSPG 783
 DB 2757 ADGRKRVQALVSHAFHSPLMDPMIDEFAVAAGAIIGRPITGIVISNTGGLADGDFSSA 2816
 QY 784 YWVRHAREVVRFPADGVVLAHAGAGTVEVGPSTLLGLVPACGPDARPALLASRRGRD 843
 DB 2817 YMRHRIRQAVRFPADSVRFQAAGSRFLVGPSSGLVASIEESLPDVAVTMTGSLRKDRP 2876
 QY 844 EPATVLEALGGLMAVGLVSMAGLFPSSGRRVPLPTYPMQRRERYI--DTRKADAA--- 897
 DB 2877 EPATLTVNAVAGFTGMDLDMRAV--GEAQFVELPTVAFQRRRPFMLSGDVAADAAGLGL 2935
 QY 898 -----RGDRRAPAGH-----DE 910
 DB 2936 AASEHALLGAVIDLPSAGGVLTGRSLSPVQGMADHSVAGVTIIFPGAQVELAIRAGDE 2995
 QY 911 V-----EEG-----GAV 917
 DB 2996 VGGGVVDESTLAAPLVLPAGGSVAVQVVNAGPDESGVRSVSRGVDGTGWLHAEGAL 3055
 QY 918 RGG----- 920
 DB 3056 RAGSASEPTADLAMPAPGAVPVEVADGYQQLAERGVGYGPAFGLTAMMRGRGDEFAEVA 3115
 QY 921 -----DR----- 922
 DB 3176 ARIAIPGSAVSIELADGLPVLISVASMALARPUTDOQLRAAVSSSGPDLFEVYMSPOP 3235
 QY 923 ----- 922
 DB 3236 SAAVEPLPVCAWGTTEDSAAVFEVSVPLAGDVVAGVYAATSSVLDVLQSWLTRDGAGLV 3295
 QY 923 -----RSARLDHP-----P 931
 DB 3296 VMTRGAVALLDGEDVDTLAGAAVWGLVRSQTEHGRITLVSDAPLDDSSALAAVVTGEP 3355

QY 932 PESGREKVEAA-----GDRPFLEIDEQVLDHLVLR-VTERRAE-GLG 974
 DB 3356 QVLMRGEVYTVARVHOSRAVGGILVPSDRPRLAMSTAGTENTLELIPADADLGGP 3415
 QY 975 EYEIADAAAGLSFNDVOLAGVPPDLPGKNPPLLGGECAGRIYAVEGNGVLVGGP 1034
 DB 3416 QVRAVASIAANFRDWMIALGLYPP-----PDAMVGEACGVVETLSINKSPAVGDR 3468
 QY 1035 VIALSAGAPATVHTSAALVPRPOLSAIEAAMVAVALTMYALADRIARLQPGERVLI 1094
 DB 3469 VMLPFPEGITVASTDQRLLVKVPAGWSHTAALTTSVPAIHAVALVDLAARSGGRVLI 3528
 QY 1095 HAATGVGLAAVQMAQVGAEVHATAGTEPKAYLESIGV--RYVSDSRSDREVADVAM 1152
 DB 3529 HNGTGVGMAAVALHNLGLVFPAT-SKGMKDTLRAMFPDDHISDSLSLEDFRRA 3587
 QY 1153 TGGEGVDVVLNLSIGELIKSFNLRSRGRFVELGKDCYADNQLGR-PFLRNLSFSLV 1211
 DB 3588 TGGRGFDVLDLSLAGEFVDSRLVAPGVFLEMGRKTDIRDQVIAQQVPGVRYRAFDFL 3647
 QY 1212 DLRGWMLERPARVALLEELGLIAGVFTPPPIATLPIARVADAFRSMQAQHLGLV- 1270
 DB 3648 E-----PGDRKIAQLAEIATLFGDGLRPLDPTTPDVCAPALRYISQARHTKVM 3701
 QY 1271 -----LTLGDPEVOIRIPTHAGA-----GP-STGDRDLDRLASA 1304
 DB 3702 LMGSGMACTVLTGTGTGMAGSAVANHVVABRGVRLVLSRRGPPAPGAELVAELAA 3761
 QY 1305 APPARAALAEFLRTVOVCYLTPKPEIK-----VGA-EALFRLGMSLMAVELRN 1353
 DB 3762 GAQVQVACDADRALAVIADIPVQHPLSGVIHTGALDADVAVMSLTPDRDVV-LRS 3820
 QY 1354 RIEASLKLSTFTFSTSPNIAL-----LAQNLUDALA 1386
 DB 3821 KVDAAHMLHELTLDLVSAFVMPSSMAGLVGSSGQANVYAANSFLDALA 3869

RESULT 2
 G86925
 probable polyketide synthase (imported) - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #ext_change 24-May-2001
 C:Accession: G86925
 R: Cole, S.T.; Eigmeier, K.; Parhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ham, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A: Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S
 A: Title: Massive gene decay in the leprosy bacillus.
 A: Reference number: A86909; PMID:21128732; PMID:11234002
 A: Accession: G86925
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-2103 <STC>
 A: Cross-references: GB:AL450380; NID:g13092513; PIDN:CAC29643.1; GSPDB:GN00147
 C: Genes: M0135
 C: Superfamily: mycocerosic acid synthase; 3-oxoacyl-(acyl-carrier-protein) synthase I h
 nase homology; lacyl-carrier-protein S-malonyltransferase homology
 C: Keywords: carrier protein

Query Match 31.9%; Score 2298.5; DB 2; Length 2103;
 Best Local Similarity 32.9%; Pred. No. 4.5e-116;
 Matches 627; Conservative 204; Mismatches 536; Indels 541; Gaps 38;

QY 4 RPIERAADPIALVGASCLPGGVIDLGSFWTLBESSRDTGVRPAER-WDAAMFDPDP 62
 DB 34 REYEQDAPTEPVAVNGICRFPGGVDSBDRLLMDVVSSECRDVSEFPDRGMDVDGLDPDP 93
 QY 63 DAPGKTPVTRASTLSVACFDASFFGISPREALRMPARILLVEWELNNAIAPSL 122
 DB 94 DAGKTYTRWGAFLDADAGDFGFIAPSEVLAMDPOORLMLVSWALEVAGIDPLSL 153

QY 123 VETGTFVIGIGPSEYEALPOATSAEIDAHOGLGTMPBSVGARI SVALGLRGPCVAVD 182
 DB 154 RSAGATVFTGIRFAPSY-----GSRDTGGLQGYGLTGTAVSVASGRVSYVLGLEGPVSD 208
 QY 183 TAYSSIVAVHLACQSLRSGECSTALAGVSLMLSTSTLYMLSKTRALAPDRGCRFASE 242
 DB 209 TACSSSLVAIHWAMSSLRSGECDLALAGVTVGPIFVGFSGRGLADRCRAFAPA 268
 QY 243 ADGPRGEGCAVVLKRLSGARADGRIIATVIRGSAIINHGASSGLGTVPNGSOEIVLKR 302
 DB 269 AGCTGMBGAGVTVLERLSDAQLGHTTVLAVVRSANVODGASNGITANGLAQORVITA 328
 QY 303 ALADGCAASVGYVEAHGTGTGLDPIEIOALNAVYGLGRVATPELLIGSVKYLNGPE 362
 DB 329 ALTNAGLSADVDVVEAHGTATTLGDPLEAOALLATVYGGGRPADQGLWNGSIKSNMGHTQ 388
 QY 363 VASGITGLKXVLSLOHGOIPAHIAQALNPRI SMQDLTLV-TRAKTWPMPMT---PR 418
 DB 389 AAGAAGAVIKVQAMKHHVMPATLHVDESPRVDTTGAVSVLTETAR---DMSVEGRER 444
 QY 419 RAGVSSFGSGTNAHVLEEAR-----AATCTPPAPERPAELLVLSARTASALDAQARLR 474
 DB 445 RAGVSSFGISGTNAHVLEEARPVEEASTADQORL SVPMVVSGRSTALMAQSKRLA 504
 QY 475 DHLETPPSQCLDVAFSLATTRSAMEHRLAVATSRERGLRAALDAAQOCOTSPGAVRSIA 534
 DB 505 FVVQADPEVDVPGCALA-ARSVEFHRVAVVGESEEQIAGLAGLAVESGAGVALIGQA 563
 QY 535 DSSRGKIALFTGQGAQTIGMGRGLYDVMSAFREARDLCVRLFNQELDPLLEVMMAEPA 594
 DB 564 -APLGKTVVFPQGGQORIGMGRGLCSQLPVFAEEFDATGELDRIMRLPLRDVVM---- 618
 QY 595 SYDPAALLDOTAFLTOPALFTEFEYALALMRSWGEPELVAGHSIGELVVAACVGVSTLEDA 654
 DB 619 GADAGILDBTEBQAPLFAVEVALFVLOHMGLODPFMVGHSIGELAAVYVAGVLAPADA 678
 QY 655 VELVAARGELMOALPAGGAMVSI EAEPADVAAVAEPHAASVSI AAVNAPDOVVIAAGACP 714
 DB 679 AMLVVARGLMQSMRPAQGTMMWAAVASEDEVSL---GEGVGIAMINAVESVVISQPOA 735
 QY 715 VVAIAAAMAAAGARTYALVSHAFFSHPLAEMLEAGRAVESYTRRPSIVLVSNLSGCA 774
 DB 736 VSAIDKFAQGRKRVHQLAVSHAFFSHPLMEPEEVRVAAQVEYKXKPOIGLVSNVTGTL 795
 QY 775 CTDEVSSPGYWRHAREVVRPADGVKALHAAGAGTVEVGPXSTLIGLPACMPDARPAL 834
 DB 796 ACPDGGSPQYMGEBHVSRAVRFVDSARHLQTLGATHITESGSSGLMGALEQSLAPAEAVV 855
 QY 835 LASSRAGDEPATVLEALGLMAVGLVSMAGLFP-SGGRRVPLPTYPMQRERYM----- 888
 DB 856 VSMWGDREBELASVILGAAGOLFATGMPEMPAVFASSSRRTVLTPTAFQRRRFMTPTGT 915
 QY 889 ----- 888
 DB 916 DGTASVSGLGISTEHALIGAVVERPDSSGVLTGRLSIANPWLADHVIGGVLPFGAG 975
 QY 889 -----IDTKADDA-----ARGDR- 901
 DB 976 FVELVIRAGDEVGCAVEBELVLAAPLVLHPGSGVOVQVVAADDESSRAVSVYSKGDOS 1035
 QY 902 ----- 901
 DB 1036 HGGWLLNAEGLELVDAEATVDSLMPPEGASVSDISDAV POLAERGVAYGPAFQGLVAI 1095
 QY 902 ----- 901
 DB 1096 WRGSELPFAEVVAPTVGVVDGGMHPALLDAVLHALGAIENQOATTEMRLPFCMFGVSL 1155
 QY 902 -----RARGAGHDEV-----EEGAV- 917
 DB 1156 HAGGAGRVARFTSAGADAIANDIADISAGLPVLTVTSVTRAMTAQLTAVTAAGGAL 1215
 QY 918 -----RGDR----- 922

DB 1216 QGPMDLVMSPIPLSSDHGGGLPSVSVSWADYCACGRNDGTGVVMEFGSASAGAQAPVDSV 1275
 QY 923 -----RSARLDH 929
 DB 1276 AAVYSATIAALQVIOFMFGQDRAGTLVLTNGAVGLVEBGISDILAAAVWGMKVRSAQSBH 1335
 QY 930 PEPESG-----RREKEVA-----GDRPRL 950
 DB 1336 -----SGKVLIDTDSA DVAVLADLGSQSLVRCNTVHAARLSPAPQLLAPVGSMMRL 1391
 QY 951 EIDPEGVLDHLYLR-VTERRAP-GLGEVEIAVDAAGLSFNDVQALGMVPPDLPGKNPP 1008
 DB 1392 AADGGCTLEDLVITQCPPEVOAPLQAGHNRVAVAAVGFVRDVAALGMVPGCAPP----- 1446
 QY 1009 LLLGECAGRIYAVEGVNGLVVGGPVALSAGVATHTVTSAAVLVLRPQALSAIEAA 1068
 DB 1447 --LGAEAGVVAIEIPETVGA VGVDAVNGFLGACAPLAV-VHQQLITQMPQMSLAEPAA 1503
 QY 1069 MPVAYLTAMVALDIRIARLQPCERRVLIHAATGCGVGLAAVQMAQVGAEVHATAGTEPKAY 1128
 DB 1504 VPVVFLLTLFGLADLAGIRAGESVLIHAGTGVGMAAQVPAHMKVEIFVTA-SRGKMDT 1562
 QY 1129 LESLGV--RYVSDSRDFVADVRAWTGSEGVVDVVLNSLSELIDKSFNLRSHGRFVEL 1186
 DB 1563 LRDMPFDDHDIGDSRTLEFEEKEFLAVTDGRGVVDVLDIADGFDVASLRLVRCGRPLEM 1622
 QY 1187 GKRDCYANQGLR-PFLARNSFSLVDLRGMKMLERPAPVRLLEBLGLIAAGVFTPPPI 1245
 DB 1623 GKTDRDAOEIAANYPGQYRAFJLSE-----AGPVMOEMLCEVRELFTQVLIHRLVV 1676
 QY 1246 ATLPIARVADAFRSMAQOHLGKLVLT-----GDPEVQIRIPHAGA----- 1288
 DB 1677 TSWDRCARPAAFRFMSQARHICKVLTMPALADLAGTVITITASAGLIGVLA RHMS 1736
 QY 1289 -----GPSTGR-----DLDRLASAAPADAALAE---FLRTQVSQVLT-PE 1329
 DB 1737 AVGVNHLVLA SRGDRAGAVELAEVLGAGAKIQUVACDVANDVADVGLFQOLSRFPFP 1796
 QY 1330 IK-----VGAELFTFLGMSLMAVELRRIRIESLTKUSTTFLSTSPNALLAQUNL 1382
 DB 1797 VGVTHAAGVLDGDAITSLTPDRVDTV-LRAVVDAMVNLHEXTR-----DIDLIMFVLC 1849
 QY 1383 DALATALS-----LERVAENLRAGVNDPVSSGADQDWE 1417
 DB 1850 SSIATVSSPGQGNYSAAANAPLDGLAAVROAAGLAGTSLANGL---WE 1894

RESULT 3

AH2140
 polyketide synthase [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:date: 14-Dec-2001 #sequence_rev1sion 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AH2140
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
 A:reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2140
 A:status: preliminary
 A:molecule type: DNA
 A:Residues: 1-2478 <XLR>
 A:Cross-References: GB:BA000019; PTDN:BAB74378.1; PTD:g17131772; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics:
 A:Gene: alr2679

Query Match 30.7%; Score 2211.5; DB 2; Length 2478;
 Best Local Similarity 25.1%; Pred. No. 2.8e-111;
 Matches 609; Conservative 242; Mismatches 502; Indels 1075; Gaps 28;


```

Db      2187  FLAGIHAGVLDGALLQNMWERQKMAKWGANMLHTQLPLDFVFLSSATL  2246
QY      1275  ----- 1274
Db      2247  FSGPGQANHVAANTFLDTLAHYRHTKGLPAMSIINWISGISATQATKMOQRGIEAIA  2306
QY      1275  -----PTHAGAGSTGD-----RDL 1297
Db      2307  PADGIQLIQLMTQPLTQIGVVPINWOKFPIINSPPFNFTLN--NDETQOTPRPKTDL  2363
QY      1298  LDRLASAAPARAALAEAFRTQVSQVLRTEIKVGALEALFTRIGMSLMAVELRNRIEA  1357
Db      2364  LEKLSQDKQGTVELLEFYLTEQIAKVLGPPRHEINPQTGEFFDIDGMSDLTAEKRNRLQT  2423
QY      1358  SLKLKSTFLSTSPNILLAQNLIDL 1385
Db      2424  DKITLPSSTIADFYPNIGTLLAHNLNQL 2451

```

RESULT 4
 H70984
 probable polyketide synthase - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: H70984
 R:Cole, S.T.; Broesh, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; PMID:98295987; PMID:9634230
 A:Accession: H70984
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1602 <COL>
 A:Cross-references: GB:Z5617; GB:AL123456; NID:g3242249; PIDN:CAM09098.1; PID:e317262;
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: pks8
 C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
 F:55-455/Domain: 3-oxoacyl-[acyl-carrier-protein] S-malonyltransferase homology <OAS>
 F:561-839/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 Query Match 30.6%; Score 2207; DB 2; Length 1602;
 Best Local Similarity 35.0%; Pred. No. 2.8e-11;
 Matches 555; Conservative 141; Mismatches 460; Indels 428; Gaps 17;

	Query Match	30.6%	Score 2207;	DB 2;	Length 1602;
	Best Local Similarity	35.0%;	Pred. No.2.8e-11;		
	Matches	555;	Conservative 141;	Mismatches 460;	Indels 428; Gaps 17;
OY	6	I ERAAEDDIAIVGASCRLLPGVVIDLSCFWTLLEGSRDVGVRPAER-WDAAMWDDPPDA	64		
Dd	28	LEAKISEEVAAYVMGCRCRYRGVDSPETIMELVACGRDAVSFPADRGNDVGLFDPPBDA	87		
OY	65	PKPTPVTRASTLDVACFDASFPGISPREALRMDPAHRLILEVCWEALENNAIAPSAVLG	124		
Dd	88	CGKMYTRGFTEFLENHADDFGFFGIGPSEALAMPQOQLLEVEWEALERTGIPOTKLRG	147		
OY	125	TETGFIIGPSEYEAAIPQATASAEIDAHGDTGMPEVSGRSIYSYLGRPCVANDTA	184		
Dd	148	SATGVAFAGVIHAGGGOL-----SCELEGYLTGSTSVASGRAVYVLTGLEPNAVSDTA	202		
OY	185	YSSSLVAHLAACQSIRSGECSTALAGVSLMSPTLVMLSKTRLARADRCKAFSAEAD	244		
Dd	203	CSSSIVALLHALAVQSLRSGECDLALAGVTWATPAAFVFPFRQALARDGRCKYYAGAD	262		
OY	245	GFGEGCAVVVVKRLSGRADDGRIILAVITGSAINHDGAASGLTVPRGSGQELVLRKL	304		
Dd	263	GTAWSEGAGVIVEELVDPARLGHPVALVIRGSAVNODGASNGILTAPMGPSQQQEVIRPAL	322		
OY	305	ADACGAASVGVVEAHGTGTLGPDIETIQALNAVYGLORDVATPLTGISVTINLGHPDEYA	364		
Dd	323	ASATLRRAVEVDVEBGHTGTMIGDPITEAOALLATT--GQDRVEPHLMGDSITSNTHGTSAA	380		
OY	365	SGITGELLKVLSLOHGQIPAHLMHAQNALNPRIISWGDLRTLVTRARPMPWDMTTPRRAGVS	424		

Db	361	AGVAGVIMVQAMHGVNPKTLHVDVPPPHVDMVSGAVSLTQPRAMVSHGPRRAGVSS	440
Qy	425	FGMSGTNAHVLEEARPAATCTPPAPER-----ABELVLSASTASALDQAARLRDHELY	480
Db	441	FGISGTNAHVLEQAPVESVVEPEASPTASAPWVWLSASSEQALQAQRLLAFLVAAN	500
Qy	481	PSQCGDVAFSLATTRSAMEHRLVAATRSRGRLDALDAAOGOTSPCAVRSIADSRGK	540
Db	501	PDLDPIDGMVLVYTRAFMEHRAVVGADRGALLAGLALAAEGSGAVVAGCRA-RSYGK	559
Qy	541	LAPLFTGCGATTCMGRLYDVWMSAFRPAFLCYRLFNQELDRPREVMAEPASVDAL	600
Db	560	TVFFPFGGAGVWGKQADLYAELPLFALAFVAAEELDRHLRLPLRNVLW-----EGDAL	615
Qy	601	LDQTAFTTPALFTTEYALAAIMRSRGVPEPLVAGHSIGELVAACVAGVFSLEDVFLVAA	660
Db	616	LTSIEFPQAPALFALEVVALATLLOHMGISPDPLIGHSGVEIAAAHLAGVLSLTDAGVLVA	675
Qy	661	RGRIMQALPAGGAVNSIEAPBADVAAAAPHAASVSIAAVNAPDQVITAGAGQPVHAI1AA	720
Db	676	RGRIMAEPLPAGGVWVWVAASEEVLPLVLDGA---NLAANVAPHSVWVSGCEAAVSIAD	732
Qy	721	AMARGARTKALNHSAPHSPMLAMPLEAFGRVAVESVSYRRPSIYLVNSLSGKACTDEVS	780
Db	733	HPARGGRVHRLVAASHAFSLMEPMLEFTRIAAGISVSKRIPLVSNNVQWMAAGYG	792
Qy	781	SPGVWVRARERVRFADGVKALHAAGACTFVVEVGPKSTLLGIVPACMDARPALIASRA	840
Db	793	DCQVWEIARRPVAFABEVQULNNAVGARTRFEVGPBGGLTALVBSLSLPGELSVAMMR	852
Qy	841	GRDEPATVLEALGGIMAVGLSVSAGLEPPSGGRVRLPLTPYMOBERVYID-----	890
Db	853	EHPVSVSLVAGVATLFTAGAQMDPVAVGSGPGRRIELPTYAFQRRGYVLPTSAGSADIS	912
Qy	891	-----	890
Db	913	GVGLLAARHGILGAVVEQPDSDVVVLTGRLSVGEQRMADHVIAGVLLAGAAFVELALR	972
Qy	891	-----	890
Db	973	AADQVDCGVBEELTVVTPVLPPTYGVQLOVVVGEMQRPVSTIYSRNAESDGVLHA	1033
Qy	891	-----	890
Db	1033	RGVLGAKVAPAADLSVMPPLGAAPVVDVDAVQRFABELGYEYGRAFGQLTMMMRRESLF	1092
Qy	891	-----	890
Db	1093	ADVAVPDDVDVTLSGFGIHPVLVDALHAMGVGEQAATMLPFSQGVSLHAAGASRVRA	1152
Qy	891	-----	899
Db	1153	RIAPAGDGTVSEVLAADQAGLPVLSVQALVMRSVSSQLLSAAVAAAADAAGRLLEVAMLPV	1213
Qy	900	-----	909
Db	1213	ELANDISADLVWELBESFQDQVGRPVYSATHRVVLALDSWLAQBRAGLVVLVTGSGVOD	1272
Qy	910	EVEREGAVRGDRRSARLDHPPPESGRREKVEAAGD-----	945
Db	1273	ATNLGAAVWGLVLSAQAEHP-----GRVMLVDSGSDMDVGVICGGEBOQLMIRNGTAYAA	1322
Qy	946	-----	988
Db	1329	RLAQLRPPPIQLPDTNSGMRLVAGAGALEBDLTLASCAPAEELAPGVQRIEVRRLGVFR	1386
Qy	989	DVQALGAVNPDDLPGKFPRLLLGCECGAGRIVAVAGEGVNGLVWGQPVATLSAGAFAFATHY	1048
Db	1389	DVLVALGIVPBAE-----LQABGAGVTEVGPVYTGIAVODPWVGL-LGVAGSFAV	1439
Qy	1049	TSAAVLVPRPOLSAIEAAMPVAVYLTWVYALDRIARLOPGERVLIHAATGCVGLIAAVOW	1108

Db 1440 VDARLVKJLJPNRWPJLTDAAQVPPVFLTAAYVALRVLAQVOPGESVLVHAAGVMAVQL 1499

Qy 1109 AAGVGAEVHATAGTPEKRAVLESIGV--RYVDSRSRDFVADYRAWVGSGVDVYNLS 1166

Db 1500 ARLMGLEVFATA-SRGMWDTLHTMGCDNTHVADSRTLAETFMLTTEGSGVDVYNLSLA 1558

Qy 1167 GELIDKSFNLRSHGRFVELGKRD 1190

Db 1559 GEFTDASLRLLPRGGRLEWKTKE 1582

RESULT 5

A12140
polyketide synthase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: A12140
R:Kanehisa, T.; Nakamura, Y.; Ujima, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.
DNA Ref. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2518 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA74379.1; PID:gi17131773; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2680

Query Match 29.6%; Score 2136; DB 2; Length 2518;
Best Local Similarity 24.7%; Pred. No. 3.5e-107;
Matches 622; Conservative 246; Mismatches 515; Indels 1132; Gaps 38;

Qy 8 RAADPPIAVGASCRPLGVIDLSEFWTLGSRDITGVPAERMDAAAFDDPDPAKG 67

Db 24 KAQTEPIAIVGMGCRFPADTPKMLQRLRNGDDGITEIPSDRWSIETFYHNNPATRGK 83

Qy 68 TPVTRASFSLDVACFDASFGISPREALRNDPAHRLLEVCWEALNMAALFSLVTER 127

Db 84 MSTRYGGFIGHLKEFDADFGIAPKEAISLDPQRLLEVTWALENTGILPRLTSSQT 143

Qy 128 GVFIIGPSEYBAALPATASAEIDAHGSGTSPVAGRIAYALGKGPVAVDTYSS 187

Db 144 GVFIIGSSNDYTHLNRPVY-DIDAYLATGNSHSTAGRISTYLGFTGSLAVDTACSS 202

Qy 188 SLVAVHLACSLRSSEGSTALAGVSLMLSPSTLVMLSKTRALARDRCRAFAEADFG 247

Db 203 SLVAVHLACSLRSSEGSTALAGVSLMLSPSTLVMLSKTRALARDRCRAFAEADFG 262

Qy 248 RGGCAVAVLKRSLGADADGRIILAVTRGSAINHAGSSGLTVNGSSGQELVLRALADA 307

Db 263 RGGCGVAVLKRSLGADADGRIILAVTRGSAINHAGSSGLTVNGSSGQELVLRALADA 322

Qy 308 GCAASSGVYBAHGTGTLGGPIEIOALNAVYGGRVATPILIGVKTMLGHEVYASGI 367

Db 323 TVEBSDDTYVEAHGTGTLGGPIEIOALNAVYGGRVATPILIGVKTMLGHEVYASGI 381

Qy 368 TGLKLVLSLQHGQIPAHHLAQAALNPRIISWGDLLTVTRATPWPMDNTPRAGVSSFGM 427

Db 382 AGLMKVVLALKHGIRPHLHFQGNPHIPWQPIQPTCTP--WHGKTLAVSSFGF 438

Qy 428 SGTHAVVLEAPATCTPP-APERPALLVLSAKTSALDAQAARLDHLETPSCCLG 486

Db 439 SGTHAVVLEAPATCTPP-APERPALLVLSAKTSALDAQAARLDHLETPSCCLG 498

Qy 487 DVAFSLATTSAA-----MEHRLAVAA-----TSRGLRAALDAAGOGSPGAVRSIA 534

Db 499 DICTSANTGSSRGAASLTQPEHRLSTIANSITTAQELTCFINNEATITNLPTNKVNS-- 556

Qy 535 DSSRGKLAFLFTQGOAGTQAGLGVWSAFREAFDLCVRLF--NOELDRPLREVMMAE 592

Db 557 -PSRPKIAELFTGQSGSQYQNMGMWELQTEPFKEETIDKCCCELLANSGLD--LHSVLFQ 613

Qy 593 PASVDAALDQAFPTQPALFTPEYALALMRSGVPELVAGHSIGELVAACVAGFSL 652

Db 614 DNEENKHLISQYLYTPALFTLEALCOLMISLGIRPDMGHSVGEYVAACLAGFSL 673

Qy 653 DAVFLVAARGRLMOALPAGGAMVSIAPADVAALVAAPASVSIANAAPDQVVIAGAG 712

Db 674 DGLKLTATBARLMQQLPQ-GKRVAVASPSQQLNPLFPDQGVSIANAAPNNTVISGE 732

Qy 713 QPVHAIAAAMAAAGATKALHSHAFSPALMELAFGRVAVESVSRPSVLYVNLG 772

Db 733 AALEKIIAVLVSONIQATPLSVSHAFSPMETMLGEFEKIAATINHPKYPPIISNVG 792

Qy 773 KACTDVSSPGVYVRAAREVVRADGVKALHAAGATFVEVGRKSTLLGLVRCMDAPR 832

Db 793 SLINSLATPEVYCRIRQVQPLAGVETLIQONCSIFLEVGAKPILLOMANSISQADK 852

Qy 833 AL-LASSRAGRDEPAPATVLEALGSLAVGSLVAGLFPSS-GGRVRLPTYPWQREYWD 890

Db 853 YMLPRLQKQDQWQVMSLSIAIYGRGVKIDHMKFEQYVHHQRDLDPVYFQRQFWID 912

Qy 891 TK----- 892

Db 913 IKPVNKVYSSSPNNIHPILGQQLNLAKSSDIYFEQQLTNNDYLDHKNFQIILPGA 972

Qy 893 ----- 892

Db 973 AYLEMSLAAKTIFKNNISIVETVSEFTCILTPDPNKAIQFILKNNHEFIVTSSND 1032

Qy 893 ----- 892

Db 1033 WITHATGKIKPHNNIPQPOINMGLONTFTKIIDITDFYQNLQVGEIGETFOATIH 1092

Qy 893 --ADDAAGDRAP----- 904

Db 1093 WYLDNOLAEIHLPSTCNCRDNYQFPIILDACTIAAIFYSQPTPSNSSODIENPSN 1152

Qy 905 ----- 904

Db 1153 LSKRRBALNPSPVKGVRGLGNTTNQVYLPFGVDKLTLYLPIGETWLSVKLRQ 1212

Qy 905 ----- 904

Db 1213 KTAFLVADIQILSPSGVIALVEGLQKKIQSPNPKDMQNMLEYEIMRSQPLSITSD 1272

Qy 905 ----- 904

Db 1273 LTPSAISOELAKFTELLTEBEIKTYAQLLPQLEALSUYIVQALQNLVLSPEIAP 1332

Qy 905 ----- 904

Db 1333 QRLYHLISLTNOREYKRPQLSPQHEIABELTIERCGTNLKVLOCKKCNPIDLFP 1392

Qy 905 ----- 907

Db 1393 NGDLTTLTQYXNSPGAKVNTLVQAINSALKDLPRGEKVRLEIAGAGGTAVVLPQ 1452

Qy 908 --HDEVE----- 912

Db 1453 LVHQSVEYVFTDISPLAKARQOFSEYEFVSQTLNIEQPLTNODITPHSFDIVIA 1512

Qy 913 ----- 919

Db 1513 LHATEMLNTTNVKSLLNQGILLVBEGLTISWIDLFGLTGEMWRFOQDLRPHHL 1572

Qy 920 ----- 919

Db 1573 ISTTAHSLKTHDPTNINIVITPDSILPRLAQGSVIVAGNTPPQSPYKREGECLIT 1632

Qy 920 ----- 919

```

Db      1633 DLPANGAALNQQKLPSTFLKPSDNIAPFPHOSIKHIIYIACQDNITECNNILHLVQTL 1692
Qy      920 -----GDRRSALDHP----- 930
Db      1693 IKTOHYPIINLWLTGCAISPHITITGLNOSVWGMKVILFELPBLNCRVDLDPTQOELFT 1752
Qy      931 -----PPESGR-----REKVEAADRPFRLEIDEPVLDHLVLRV 965
Db      1753 QVENLVTEITHPGEAEQITLHKSERKARLTCTQLILPHEPRLTIARGVASLKMQS 1812
Qy      966 TERRAPGGEVEIADNAGLSFNDVQLAGM/PDDLPGKPNPRLLLGGCAGRIYAVGEG 1025
Db      1813 SRRRPPQPEVEIIOYATGLNFIIDVLDGL---LPFRN---WFGVCAEVAVALIGG 1865
Qy      1026 VNGLVGQFVILASGAFATHTTSNALVPPRQALSAIEAAMPAYVITAMVALDRIR 1085
Db      1866 VTHLVGDAVVALADSPQYVTTNANYIKKPDLSFTMAATITANFILTAAVLEKAVK 1925
Qy      1086 LQGERVLIHAATGGVGLAAVQMAOHVGAETHATGTPKRAYLSLGVRYVSDRSDFR 1145
Db      1926 IQPKRIILHAATGGGMAALQIAQOAGLEVFATASV-GKMETLRALGVQHI.FNSRTLNF 1984
Qy      1146 VADVAAAMGEGVDVNLNSLSEGLDKSFNLRSHRFEVLGKDCY-ADNQLGRPFGR 1204
Db      1985 AEHIEITQEGVDIVFNSLSEGFIPASLSVLKPGQFLEIGKGVMAQKVAQVFP--- 2041
Qy      1205 NLSFSLVDLRGMLEPRPARVRALBELGLLAAGVF-----TPPIATLPIARVADA 1256
Db      2042 DVYVHLVDLMSVAQOQFOTITQTLHCLMAEFESGELRMSPTTRSSHPTITPATKIVEA 2101
Qy      1257 FRSMAOAHGLKLVLT----- 1272
Db      2102 LQMMQOARHIGKIVITHTPTESLQPDATVYLTGKMGGLGARVAVLVEKAGHVLVLGRN 2161
Qy      1273 ----- 1272
Db      2162 QPDTTAQOQIALGATGAKIITTOADVSQKQDLAAVLTDIQONHPRLRGVTHAAGVLDDG 2221
Qy      1273 -----LGDP----- 1277
Db      2222 ILQQLTPERSHRVMPKVTGAMNLIQLODITLDFIMSSSAALSGSGQANHYAANTF 2281
Qy      1278 ---VQIR---IP-----THAGAPSTGDR-DLID--- 1299
Db      2282 LDALQVRRHQGLPALSTINMGVMSDIGAAKQVSNQMSRGIGETTPQGDILIEFLT 2341
Qy      1300 -----RLAS-----AAPA-----RAAA 1312
Db      2342 QSTGVGVPIINWSELKQKLSSTFPADFEYESTPSTOEYQPSOLIQHIQCLGKERISY 2401
Qy      1313 LEAFRTQVSOVL-----RTPEIKVGAELFTRLGWDLSMAVELNRRIREASLKLSTTF 1367
Db      2402 LKRIQTEVSQVLGSSSSQLPWNQUG---PDMGWDLSLMWELSLRSLTSKQITPSTV 2457
Qy      1368 LSTSPENLALQNLDAATATLSLR--VAAENLRAGVQNDPVSSGADQWEIIA 1420
Db      2458 LPEHRSI-----QALAEVIATQLLPQETVISAINEQPDISSDVSSKQEIETSI 2508

```

RESULT 6

AB2012
 Hypochemical protein all1648 [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp. PCC 7120
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
 C:Accession: AB2012
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Itiguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AB2012
 A:Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-1587 <KOR>
A:Cross-References: GB:BA000019; PIDN:BA878014.1; PID:G17135468; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all1648

Query Match      29.4%; Score 2120; DB 2; Length 1587;
Query Similarity 44.7%; Pred. No. 1,4e-106;
Matches 450; Conservative 168; Mismatches 337; Indels 52; Gaps 14;

Qy      6 1ERRAEDPIAIVGASCRLLPGVIDISGFTLLESGRDTGVRVPAERMDAAWFDPPAP 65
Db      26 LERTQNEBIAIVGMCRRPPGDANNPEKEMELLRQCKDITTPPQRMIDIDAYDDDDPPV 85
Qy      66 KCTPTTRASFLSDVACFPASFFGISPRALMDPANHLLLEVCWALNNAAIAPBALVGT 125
Db      86 NKMVARYGCFINNVQDFDPQFEGITPREAIALDPQRLLEVSMALENAGIAPQKLTGT 145
Qy      126 ETGVFIGIGPSEYEALPQATASAEIDAHGIGTWPVSGARISYALGIRGCVAVDTAY 185
Db      146 QTVGVFGIGIDDY--AKQIKHHPIDAYTSGNACFCAAGRLSTLGLQPSLAIDTAC 203
Qy      186 SSSIVAVHACQSLRSGECSTALAGVSLMLSPSTLWLSKTRALARDGRCAFSADAG 245
Db      204 STSLVTHLACQSLRNGECNLALAGVSLMLSPSTLWLSKTRALSPDGRCKTFORDANG 263
Qy      246 FGRGEGCAVVVLKRLSGARADGRILAVIRGSAINHDGSSGLTYPNSSQGEIVLKRALA 305
Db      264 YVRGEGCGMNVVLKRLSSAVADGDHLAVIRGSANVQDASSGLTYPNSTAOADVROLA 323
Qy      306 DAGCASSGVYEAHGTGTTGDPFIEIQALANAVVGLGRDVAETPLLIGSVKTNLGHPEYAS 365
Db      324 NAKTTPAOTSTLEAHGHTALGDPLEVAIADIGVRGKGSNNHPLIGSVKTNIGHLETA 383
Qy      366 GITGLKVVLSLQHQIPAHLLHAQALNPRISMGDLRLTVTRARPPWDMNTPRRAGVSF 425
Db      384 GMASTLKVYLSLQHQEIPPHINPQELNPDLLAASAKSLKIPISYIMQCTEGRMAGISSF 443
Qy      426 GMSCTNAHVLEEARPACTPPAPERPARELVLSKRTSALDQAQARLDHLETPSQCL 485
Db      444 GLSGTNNAHIIIEEPQQLVTPEAVDRPLHVLMLSKSAPALHTLATDMEHLRNHPETNF 503
Qy      466 GDVAFSLATTSAMEHRLAVALAATSRREGRLALDAAQOCTSPGAVRSIADSSR-GKLAFL 544
Db      504 ADLAFSANTGSGFNHRLAIIAQSTQAKQVLAENQQLPBLNPFSSQVEVGKQCKIAFL 563
Qy      545 FTGQGAQTLGNGRGLYDVMSAFREAFDLCVRLFNQELDRPLREVMMAEPASVDALLDQT 604
Db      564 FTGQGSQYVGMGRQLYETQPTFRQALDECDRLDQYLKESLSLVLYPQTPPAN-PLINQT 622
Qy      605 APTQALFTPEYALALMRSVNGVEBELVAGSISGELVACVAGFSLDANFLVAARGRL 664
Db      623 AYTOTALPAIETAYALCKLMQSWGIOPOGVLGHSVEBYVAACTAGVYTLQEGIELIAQORQL 682
Qy      665 MOALPAGGAMVSIAPREADVAAVAAPHASVISAANVAPDOVYIAGAGQPHAIAPAAAMA 724
Db      683 MOALPQSTMAAVPAPVETVARAIAPYANETITATINSPEVNVISGVAAIALVADULTA 742
Qy      725 RGARTKALVSHAFSPILMAAPMLEAFGRVAESVYRPSIYLVSNLSKACTDEVSSPGY 784
Db      743 QGIDVRLQVSHAFHSPMEPMLEFQVAAKINQOTGIDMISVTAETHSIDAE-Y 801
Qy      785 VYRHAAREVRRADGVKALHAAGAGTGVVGVKSTLGLVPCMPDARALLAASSRAGDE 844
Db      802 WCOQIRDOVQAPAMETLAAQGVYVLEIGHVPLVTRGKQTLSDPPQILMLPSLHRENN 861
Qy      845 PATVLEALGMAVAGVLSVAGLFPSPG-GRVPLPTVQWQERYWIDTKADDAARGDRA 903
Db      862 WQSLQSVATLSVGVRLDMSGFEQDYIRRLITPTTPFOQORVWL-----AAESTQ 914
Qy      904 PGAGHDEVEEGAVGGDRRSARLDHPPESGRREKVEAAGDRPFRLEIDEPVLDHLVLT 963

```

Db 915 P-----EVIPVAI-----SPET---STIVA-----TETLESQILS-LVA 946

Qy 964 RYTERARPGIGEVIAVDA---AGLSFNDV---QLMGWPPDLLPK 1004

Db 947 KIT-----GNNPQSLDALTLEGGLGDSITMTQMLNGLIKTLPQ 988

RESULT 7

S23070

erythronolide synthase (EC 2.3.1.94) II - Saccharopolyspora erythraea

N:Alternate names: 6-deoxyerythronolide B synthase II

C:Species: Saccharopolyspora erythraea

C:Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text_change 21-Jul-2000

C:Accession: S23070; S23011; S23205

R:Bevitt, D.J.; Cortes, J.; Haydock, S.F.; Leadlay, P.F.

Eur. J. Biochem. 204, 39-49, 1992

A:Title: 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of

A:Reference number: S23070; MUID:92155230; PMID:1740151

A:Accession: S23070

A:Molecule type: DNA

A:Residues: 13573 <BEV1>

A:Cross-references: EMBL:X62569

R:Bevitt, D.J.

submitted to the EMBL Data Library, September 1991

A:Reference number: S22011

A:Accession: S22011

A:Molecule type: DNA

A:Residues: 1-184, 'T', 186-301, 'S', 303-521, 523-658, 'A', 660-993, 1001-1212, 'H', 1214-1392, 13

3479, 'DH', 3480-3572 <BEV2>

A:Cross-references: EMBL:X62569; NID:946977; PIDN:CAA44448.1; PID:G581651

R:Gaffrey, P.; Bevitt, D.J.; Staunton, J.; Leadlay, P.F.

EMBS Lett. 304, 225-228, 1992

A:Title: Identification of DEBS 1, DEBS 2, and DEBS 3, the multienzyme polypeptides of th

A:Reference number: S23103; MUID:92316235; PMID:1618327

A:Accession: S23205

A:Molecule type: protein

A:Residues: 2-12, 'XXX' <CAF>

A:Experimental source: strain CA340

C:Genetics:

A:Gene: eryA

A:Start codon: GTG

C:Function:

A:Description: catalyzes the construction of a polyketide chain, which is then cyclised

C:Superfamily: [acyl-carrier-protein] S-malonyltransferase homology; 3-oxoacyl-[acyl-car

ort-chain alcohol dehydrogenase homology

C:Keywords: acyltransferase; antibiotic biosynthesis; carrier protein; coenzyme A; multi

F:52-453/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F:561-843/Domain: [acyl-carrier-protein] S-malonyltransferase homology <SAMT1>

F:1140-1308/Domain: short-chain alcohol dehydrogenase homology <SAD2>

F:104-1475/Domain: acyl carrier protein homology <ACPI>

F:1519-1919/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

F:2023-2305/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>

F:2857-3131/Domain: long-chain alcohol dehydrogenase homology <LADH>

F:3149-3327/Domain: short-chain alcohol dehydrogenase homology <SADH>

F:3420-3493/Domain: acyl carrier protein homology <ACR2>

F:1139/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 29.0%; Score 2093.5; DB 2; Length 3573;

Best Local Similarity 32.3%; Pred. No. 1,1e-104;

Matches 607; Conservative 178; Mismatches 532; Indels 561; Gaps 45;

Qy 2 ADPRIPRAE-DPPIATGACRLPGVITDLSGFMTLL-EGSRDVGKVPAR-WDAAMF 58

Db 1487 AVRVAADSESEPTIATIGICRPGGIGSPQQLRWVLAEGANLTTG-FPADRGMDIGRLV 1545

Qy 59 DDPDARAGKTPVTRASFSLDVACFDASFPFGISPREALRMDPARHLLLEVCEALENAIA 118

Db 1546 HPPDDNNGTSTVVDKGGFLTDADFPGFGLTPREALMDQRLMETANEAEARICID 1605

Qy 119 PSALVGTETGVFIGIPSEVEEALPQATASAEIDAHGGLGTMPSVGARISYALGRLGPC 178

Db 1606 PDALRGTDIGVEVMNGGYSWOLL--AGEAEKRVDSYOGIGNSASVLSGRITAVTFEGMEGA 1663

Qy 179 VAVDPAVSSTVAVHLAAGSLRSRGCSTALAGVSLMSPTLVMLSKTRALARDGRCA 238

Db 1664 LVVDTRACSSLLVGIHLAMQALRRGCSTLACGAVTMSPTLVFVDFSTRGSLASDRCIA 1723

Qy 239 FSAADGFGRCGCAVVLKRLSGARADDRILAVIRGSAINHDGASGLTVPNGSSQEI 298

Db 1724 FSAADGFGALSQVAVLVEPIRSARANGQVLAVIRGSVAVQDASNGLAAPNGSQER 1783

Qy 299 VLKRALADGCAASSGVYEAHGTGTTCDPIEIQALNAVYGLGRDVATPLIGSVKTL 358

Db 1784 VIKQALAAAGVPAADVDVEAHGTGTECDPIEAGLLITVYQODRD--RPLIGSVKTI 1841

Qy 359 GHPEVNSGITGLKVLSTHQGQIPAHILHAQNLNPIISGDLRLVTRARTPMPDMNTR 418

Db 1842 GHTQAAAGAGVYKVLARHQLPRSLHADLSPIIDBSGAVEYLRVEVMPACERR 1901

Qy 419 RAGVSSFGSGTNAHVLEAP-----AATCTPPAPERAEILLVSARTASALDAQARL 473

Db 1902 RAGVSSFGVSGTNAHVLEAPAEQEAARTERGPIV-----FVLSGRSEAVVAQAARAL 1955

Qy 474 RDHLETYPQCCGDVAFSLATTRSAMEHRLVAATSREGLRALDPAAGQTSPPGAVRSI 533

Db 1956 AEHLRDTPELGLTDAAMTLATGRARPDVAAYVAGDDRAGVCAELDALAGRPSADVAPEV 2015

Qy 534 ADSRRGKLAFLFTGQAGQTLTGKRGLYDWSA.PREAFDLCVRLFNQELDRPREVMMAEP 593

Db 2016 TSPVPR-KPVLVPPGQCAQWGMARLDLSSSEVFASMSCAEALSHHTTKLLDVVRGCG 2074

Qy 594 ASYDALDLDTQTAFTQPALTFEYALAAALRSWGVBEYLAHSGISGLVAACVAGVSLSD 653

Db 2075 GPDHPRVD---VLQVLFVSIWLSLELRAHGVTPAAVVGSHQGEITAAHVAAGALSLBA 2131

Qy 654 ANFLVAARGLMQLPAGAMVSIPEAPDVAAVAAPHAASVIAAVNAPDQVITAGAO 713

Db 2132 AAKVVALRSQVRELDDQGMVSVGASRDELTVALRMQDRAVAAVNAPGTSVAGPFA 2191

Qy 714 PVYALIAAMAAGARTKALHVSAPSLMAPMLTFAFGRAVSVSYRPSIVLVSLSGK 773

Db 2192 ELDEFFAEABEMKRRIRAVRYASHSEVARIEDLAELEGTITVAVRGSVPLHSTVTGE 2251

Qy 774 ACTDEVSPPGVWVHAREVVRPADGVKALHAAGAFTVEVGPKSTLL-----GLV 823

Db 2252 VIDTSMADSVMWRNLRVLFQAVRGLVEQCFDFVEVSPHVLMAVEETAEHAGAB 2311

Qy 824 PACMPARPLLASSPAGDEPATVLEAL--GGLNAVGLVSWAGLP--SGGRKVPPLPT 879

Db 2312 VTCVP-----TLRREQSGPHIEFLRLRAHVHVG--ADLRPAVAGGRPAELPT 2358

Qy 880 YPMQREBYM----- 888

Db 2359 YPEHOREFMRPHRPADVSALGVGAEHPLLAADVPGHGAVFTRGLSTDEQPLAEH 2418

Qy 889 ----- 888

Db 2419 VVGGRTLVGSVLVDLAAAGEDVGLVLEELVLRPVLVLAGAGALLRMSVGA.PDSGRR 2478

Qy 889 -IDTKADD-----AARGDR----- 901

Db 2479 TIDHAEDVDLADAOQMSQATGTLAQVAAQGRDTEQWPPEDAVRILPLDDHYDLAGEQ 2538

Qy 902 -----RARGHDEV-----EAG-----GAV----- 917

Db 2539 YETGPEFQALRAWRKQDSVYAEVSLAADDEGIAFHPLVLDVAQTLISGALGEPGGKL 2598

Qy 918 -----RGDDRARL-----DHP 931

Db 2599 PFAMNTVTLASGATSVRVATPAGADAMALRTDPAGHLVATVDSLVVRSTGEKMQE 2658

Qy 932 PESGRRE-----KVEAGRRP-- 947

Db 2659 PRGGEGLHALDWRLAEPSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPRA 2718

QY 948 ----- 947
Db 2719 AARHGVLMAALVRRLLEQELRGATLVATSGAVTVSDDDSVBEGGAAMGVCIRCAQA 2778
QY 948 -----DHVLRV-----TERRAPG- 972
Db 2779 ESPDRFVLLDTAEFGMLPAPVPDPNQLRGDDVFVPRLSPLAPSLTEAGTQRDVPD 2838
QY 973 -----LGEVLAIVDAAGLSFNDVQLATGMVDDDLGKXNPFLILG 1012
Db 2839 GAIDVAEPAPDVEQPLRRAGSVRVADVPATGVNFDVLLATGMYPQKAD-----MG 2890
QY 1013 GECAGRIVAVGEGVNLVVGQPIVATLSAGAFATHVTSALVLRPQALSAIEAAMPYA 1072
Db 2891 TEAAGVTVAVGPPVDVAF-PGDRVLGFGQAFAPATDHRLLARVGDGSDADAAVPTA 2949
QY 1073 YLTAMVYALDIARLQGERVLIHAATGVGLAAVQAHVGAENVHATAGTPEKRAYLES 1132
Db 2950 YTTAHVYALHDLGLRAGQSVLIHAAAGVGMAAVALARABAGAVLATAG-PAKHGTLRAL 3008
QY 1133 GV--RVVSDRSRPFADVRAMTGEVUVVNLISGELLIDKSFNLRSHGRVLEGRD 1190
Db 3009 GUDDERIASSRETGFARKFRERTGKGVVNLISLTGELLIDBSADLLADGVFVEMGKT 3068
QY 1191 CYADNQLGLRPFLRLNLSFSLVDLRGMLEPRA-----RVRLLEELLGLLAAGVF 1240
Db 3069 -----LRDAG-----DRG-----RYAPFDLGEAGDDRLGSLIEVVGLLGAGEL 3108
QY 1241 TPEPIATLDIARVADAFRSMAQOHLGKLVLTG--DPEVOIRIPTHAGAGSTGRDL 1297
Db 3109 DRLPVSAMVLSGAPALQHMRSRGHVMVLTQPAVDPDGVLI-----TGCTGT 3159
QY 1298 LBRLSAARAAARAAALAEALFRQVQSVLTPREIKVGAELFTRLGMDSLMAVELRRITA 1357
Db 3160 LGRLL-----LARKLVTEHGR-----HLLVSRGADAGCSDELRAEIE- 3197
QY 1358 SIKLKLSTFTFLSTSPVIALLAQNLIDALATL-----SLERVAEN-DR 1400
Db 3198 DLGASAEIAACTADADALSA--LIDGLRPLTGVVHAAGVADGLVTSIDERAHVQVLR 3255
QY 1401 AGVQNDVSSGADQWEI 1418
Db 3256 AKV-----DAANVL 3264
RESULT 8
T17467
rifamycin polyketide synthase modules 9-10 - Amycolatopsis mediterranei
C:Species: Amycolatopsis mediterranei
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #ext_change 17-Nov-2000
C:Accession: T17467
R:Schupp, T.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z18802
A:Accession: T17467
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3413 <SC>
A:Cross-references: EMBL:AJ22012; NID:e1227119; PID:e1227124; PIDN:CA11039.1
A:Experimental source: strain LBG A3136
C:Superfamily: acyl carrier protein homology
C:Keywords: carrier protein
F11608-1679/Domain: acyl carrier protein homology <ACPI>
F13334-3405/Domain: acyl carrier protein homology <ACP2>
Query Match 28.6%; Score 2063.5; DB 2; Length 3413;
Best Local Similarity 35.1%; Pred. No. 4.3e-103;
Matches 601; Conservative 195; Mismatches 547; Indels 369; Gaps 56;
QY 8 RAADIEDIAVAGSCLPGYIDISGFWTLLEGSRDVGKVPAR-WDAAMFPDPAAG 66
Db 27 RHADIELAVVAGACRPPGVSSPEDLMOLVAGGVADLSDFPDRGWELDGLFDPDPDHG 86

QY 67 KTVTRASFSLDVACFDASFFGISPREALRMDPAHRLLEVCWEALENAAIAPSLVGT 126
Db 87 TSYTQGGFLRAGLFDAGLFCISPREALVMDPQORVLETSWEALIEDGCVPLSLKSD 146
QY 127 TGVFTIGIBSEYEALPOATASAEIADHGGCTMPSVGAGRISVALGLRGCVAVDTAYS 186
Db 147 VGVFSGVFTQGYGA----CAIPDLEAPAGIGAASSVSGRVSYVFGLEGPAVITIDTACS 202
QY 187 SSVLVNHLACGSLREGCESTALAGVLSLSTLWLSKTRALARDGCKPFSAEADPF 246
Db 203 SSVLVNHLAAQQLRAGECSMALAGATVMPGTVAASRQVLADGSKAFSSSTADT 262
QY 247 GREGCAVVLRLTSGABADGRILAVIRGSAIINHDSASSGLTVPNGSQEIVLKRALLD 306
Db 263 GMAEAGVVLVERLSVADGERHILAVIRGSANVDGASNGITANBPGQGVIRKALLAG 322
QY 307 ACCAASSGVYBAHGTGTTLGPDIETQALNAVYGLRDVATPLIGSVKTNLGHPEYASG 366
Db 323 AGLVASDVDDVBAHGTGTLGDPIDBAQLLATYGGRE--RPLMGSVXSNFGHTQAAAG 380
QY 367 ITGLLKVYLSLQNGIIPAHILHAQALNPISMGDLALTITRATPPMDMTPRAGVSSRG 426
Db 381 VAGVIAKMQALRHGAMPPTLHAAPTPEVDSAGAVELTEPERMPAGRPRAAGSARF 440
QY 427 MSGTAAHVLEAPATCTPPAPERPAEL-LVLSARTASALDAQARLRDLETYPSOCL 485
Db 441 IGTNAHLILEAPPADVAEPEPKGPVLYVVSAGSPSLAAQAGRLAEVLAS----- 494
QY 486 GDV-----AFSLATRSAMEHRLAVALTSREGRLAALDAAAGQTSPGAVSRIADSRGK 540
Db 495 GGVSBARLASGLISGRALLGDRAVAVVAGTDEDAVAGLRALARGDRAPGVLTGSA--KHGK 552
QY 541 LAFLTGCGAQTLLGNGRGYDVWSAFREAFD-LCYRLRNOELD-----RPLAEVMA 591
Db 553 VYVFPFGQSQRLGNGRELVDYRPVAFATFDAQ-----EQLDVCLAGRAGRVDVVLG 607
QY 592 E-PADVDAALDQTAFTOPALFTFEYALAAALRSWGVDEBELVAGHSIGELVAAVAGVS 650
Db 608 EYPA--ETGLNQTFTQAGLFAVESALFRLAESGVPRDVLGHSIGITAAVYAGVS 665
QY 651 LEDAVFLVAARGRLMQALPAGAMVSIAPADVAANAAPHASVSIANAAPDOVVIAG 710
Db 666 LDDAARIYAARGRLMQALPAGAMVVAASAEEVLELGD--GVELAAVNPSPAVVLSG 722
QY 711 AGCPYHALTAAMAAGARTKALHVSHTSPMAEMLEAFGRVAESVSRRRSYLVLSL 770
Db 723 DADAVVAAAARMRERGHKTQKLVSHAFISARMAEMLEFAELAGVTRREBEIPVSV 782
QY 771 SGK-ACTBESVSGVWVHAREVVRPADGVKALHAAGAGTFVEVGPCKSTLLGLVPACMD 829
Db 783 TGRFAEPGEELTPGVAEHVRPVRFAEGBVAAATSSGSLFVELRGALTLVE---ET 839
QY 830 ARPALLASRAGRDEPATVLEALGVLWVAGVLSWAGLPFSGGRVPLPTYPWQRRYWI 889
Db 840 AEVTCVAAALRDRPREVTALITAVAELEFRGVAVDMPALBPVTGFDLPKVAFDQHWYL 899
QY 890 D--TKADDA-----KGRBRARCA-----GHDEVE---GC----- 915
Db 900 OPAQATDAASLGVAADHPILLAGVAVRLPQSDGLVTSRLSLKSHPMWLDHVGIVLVA 959
QY 916 -----AVRGDRRSARL-----DH-----PPESGR-----EKV 940
Db 960 GIGVELAVRAGDEAGCPVLELVLEAPLVVDHGGVRIQVVVAGPAGETGSAVEVYSIR 1019
QY 941 EAAG-----DRPF-----RLEIDE--PGVLDH-----L 961
Db 1020 EDAGAEVNAARHATGFLAATPSQHKPFDTAMPPEGVREVDVFDGLVDGRYAGPSFR 1079
QY 962 VLVTRERAPGL-GEVEIA-----VDAAGLSFN-----DVQALGM-----VPDDLPGKPNP- 1008
Db 1080 GLRAVWRKDEVFAEVALAEDRDADARGLIHPGLLDALAHGAGAGATTTEPRGPVLPF 1139

Qy 1009 -----LLIGCEGACRI-----VAVGEVGNLV-----VGQVITALSAGFAT 1045
 Db 1140 AMNGVLHAAAGASALRLRLASGPDALSYEAADAGLVVTAADSLVSRLPSAQLGAAN 1199
 Qy 1046 H-----VTTSAALVLRPPQALSAT-----EAAAMPVAVLTAWYALDRI--ARLQP 1088
 Db 1200 HDALFRVEMTEISSAGDVPAADHVEVLEAAGEDEPLELTVGLVLEAVQWLA--DAADDAARLV 1258
 Qy 1089 GEEVLHHAATGAGVLAIVQAQHVGAHVHTAGPREKAYLESIGV-----RYSVDSRSDR 1144
 Db 1259 VTTGAHVETDPAGAAV--N-----GLTAAQENPDIIVLIDTDGEVPLGRVLTATGEPT 1312
 Qy 1145 FV-----ADVAMTGEQVDVIVNSLSEGL-IDKSFNLRSHG--REVEL 1186
 Db 1313 AVRGATLFAARLARAEAEAPAVTGG-----TVLISGAGSLGALTARHLVARHGVRLV 1368
 Qy 1187 GKRDCAADNOLGRPFRLNLSFSLVDIRGMLERPARVRLLEE-----LIGLIAAG 1238
 Db 1369 SRGPDADGMAELTAEILIAQGAEEVAVACDLADRD--QVRVILAEHRPNAVVHTAGVLDG 1427
 Qy 1239 VTPRPATLPIARVADAFR--SMAQAQHLGKL----- 1269
 Db 1428 VF-----ESLREKLAKVFAPKVTAAHNLDELRELDLRAFVVFSSASGVFSAGQNYA 1482
 Qy 1270 -----ESLREKLAKVFAPKVTAAHNLDELRELDLRAFVVFSSASGVFSAGQNYA 1277
 Db 1483 AANAVIDAVANRRRAAGLPCTSLAMGLMEQTDGNTALGDADQARASRGVLAISPAGM 1542
 Qy 1278 -----VOIRI---PTHAG-----AGP-STGDRDLDRLASAA 1305
 Db 1543 ELFDAAPDGLVVPKLDLRKTRAGCTVPHLRGLVRPGRQARPASTVDNGLAGRLAGLA 1602
 Qy 1306 PAARAAALEAFLETPVQSQV--RTRPEIKVGAELFTLGMDSLMAVLEARRIEASLKL 1363
 Db 1603 PAQOEALLDVKRFQVLLVGHAGPB--AVRADTAFKDTGPFDSLTSVELRRLRBSAGKL 1661
 Qy 1364 STTFLLSTSPNIALIACNMLDALATALSLEVA 1395
 Db 1662 PATLVFPYPTPVALLARLRLDELGDVATTTPVA 1693

RESULT 9
 117410 polyketide synthase type I - Streptomyces venezuelae
 C:Species: Streptomyces venezuelae
 C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
 C:Accession: T117410
 R:Xue, Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.
 Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
 A:Title: A gene cluster for macroide antibiotic biosynthesis in streptomyces venezuelae
 A:Reference number: Z18773; MUID:9844533; PMID:9770448
 A:Accession: T117410
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Residues: 1-3739 <XUE>
 A:Cross-references: EMBL:AF079138; NID:G3808326; PID:G380835; PIDN:AA069330.1
 C:Genetics:
 A:Gene: pKAT1
 C:Superfamily: acyl carrier protein homology
 C:Keywords: antibiotic biosynthesis; carrier protein
 F:145-1516/Domain: acyl carrier protein homology <ACPI>
 F:3570-3641/Domain: acyl carrier protein homology <ACP2>

Query Match 28.5%; Score 2052.5; DB 2; Length 3739;
 Best Local Similarity 28.0%; Pred. No. 1.96-102; Indels 811; Gaps 36;
 Matches 604; Conservative 190; Mismatches 551;

Qy 5 PIERAAD-PIAIVGASCRLLPGVITDLSGFWTLLEGSRDTPVGRVPAER-MDAAMFPDP 62
 Db 1532 PVDSGVDDPFAIVGMAICRLFGVASPDMLRLVAGGEDAISGFPQDRGMDVEGLYDPDP 1591
 Qy 63 DAPKRTVTRASFSLDVACDPASFFGISPREALMDPAHRLLEVCWEALENNAIAPSAL 122

Db 1592 DASGRTYCRAGGFLDEAGEFDADFFGISPREALAMDPOORLLLETSMWEAVDAGIDPTSL 1651
 Qy 123 VGETGVTFIGIGBSSEVEALPQATSAEIDAGGCTMPSVGAGRISTVALGRPCVAD 182
 Db 1652 QGQGVVFPAGTNGPHEPL--RNTAEDEEGVGTGNAASINGRSTYLGLEGPAVYTD 1709
 Qy 183 TAYSSLVAVHILACOSLRSGECSTALAGVSLMSPSTLVMLSKTRALARADRCASFAS 242
 Db 1710 TACSSSLVALLHIAVQLRLRGEGLAGVTVMTSTPTTVEFRSROGLAEDGRSKAFAS 1769
 Qy 243 ADGFGGECGCAVVLKRLSGARADGRILLAVIRGSLINNDGASSGITVNGSQEVLVR 302
 Db 1770 ADGFGGECGCAVVLKRLSGARADGRILLAVIRGSLINNDGASSGITVNGSQEVLVR 1829
 Qy 303 ALADGCAASSVGVYAHGCTGTLPRIETIOLANVYGLGRDPAFTLIGSVKTNHGE 362
 Db 1830 ALADGCAASSVGVYAHGCTGTLPRIETIOLANVYGLGRDPAFTLIGSVKTNHGE 1889
 Qy 363 YASGITGLKVLVSLQHQI PAHLHAQALNPRIISWGLRLTYTRATPMP--DMNTPRA 420
 Db 1890 AAGVGSIILKVMQAMHGVLPKTLHVDRPSDQIDMSAGTVELLTEAMDPKQEGGLRA 1949
 Qy 421 GVSFGHSGTNHNVLEAPATCTPPAPRABELV--LSARTSALDAQALRLDL 477
 Db 1950 AVSSFGISGTNAHIVLEAPVDE--DAPADEPSVGVPVPLVSAKTPAALDAQIGR---L 2004
 Qy 478 ETYPSOCL-----GDVSLATTRBAMEHRLVAATRSRGLRAALDAAGOTSGAVR 531
 Db 2005 AAFASGRTDADPGAVAVKLAGRQAFHRAVALTGODDLAALAA-----PGLVR 2058
 Qy 532 SIADSRGKLAFLFTGQAGTLMGRGLYVMSAFREARDLCVRLFNOELRPREVMA 591
 Db 2059 GVA--SGVGRAPVFPQGGQMGMAELLDVSEFPAAMAECAALAPVYDMSLEAVNRQ 2117
 Qy 592 EBPASVDALLDQTAFTQPALFTFEVLAALMSKSGVEPELVAGHSIGELVAACVAEVE 651
 Db 2118 APG---APFLERVDVVPVTFVAVVSLAKVQHNGVTPAVVGHSGEELAAAVAGALS 2174
 Qy 652 EDAVFLVAAAGRMOLPAG--GAMVSIAPBEADVAAVVPAHVASVIAAVNAPDVVING 710
 Db 2175 DDARVVTLSKSIKGLHAGGMLSLASEAVVRLAGF--DGLSVAAVNGTAVVSG 2233
 Qy 711 AGQVHAIAAAMARAKTKALVHSHAFHSPLMAPMLEAFGRVAESVYRPSIVLVS 770
 Db 2234 DPTQIGELAQACADGVRAIIPVDYASHSAHETIESELADVLGLSPQTPVPFFSTL 2293
 Qy 771 SGACTDEVSPPGWRHARBEVVRPADGVKALHA--AGAGTFVVGSKSTLLGLVPACMP 829
 Db 2294 EGAMITEPALDGGYWRNLRHVGFAVETLATDEGFTHFVEVSAHPVLTMLPETVVG 2353
 Qy 830 ARPALASSRAGDEBPATVLEALGMAVGLVSMAGLFPSSGRARVPLPTYPQWRERYW-- 888
 Db 2354 ---LGTLRKRDNGGGRHLTSLAEAMANGLYVDMASLPTTTTHDLPFYAQTERTYMP 2408
 Qy 889 ----- 888
 Db 2409 QPDLSSAAGDITSAGLGAHEHPLGAVALADSDCLLTGSLSLRTHPMLADHAVAGTVLL 2468
 Qy 889 -----IDTKAD----- 895
 Db 2469 PGTAFLVAFRAGDOVCDELVEELTLDAVLVPRGAVRQVLSVGSASDSGRTTGLYAH 2528
 Qy 896 -----AARGDRAP-----GAGH-- 908
 Db 2529 PEDAPGAETRTATGVLARADRTAPVADPEAMPPEGAPVVDGLYERFANGCYGCE 2588
 Qy 909 -----DEV-----EEGA-----VRGDRRS 924
 Db 2589 LFGGVGRWKRGEVEFADVALPAEVAGAGARFGHLPALDLDAVQAAGAGRGVRRGHA 2648
 Qy 925 ARLDH----- 929
 Db 2649 VRLERDLVAVGATATLVRVLAPAGPTVSVAADSSGQVFAADSLTVLPVDPQOLAFA 2708

```

QY 930 ----- 929
Db 2709 DFTLADHLEWTAMWGAAQALBGAIVVGGADAGLAAALRAGTEVLSFPDLTDLVEAND 2768
QY 930 ----PPES----- 934
Db 2769 RGETAPATVAVACPAAGPDGPEHVREALHSLALMQAMLADERFTDGLVTRDAVA 2828
QY 935 -----GREKVEA-----ADRP----- 947
Db 2829 RSGDGLRSTGQAAVWGLGRSAQTESGPFVLLDLAGEARTAGDATTGDTTGDATVGGT 2888
QY 948 ----- 947
Db 2889 SGDAALGSLATRALSGEPQALRDGALLVPRILARAAPAAADGLAAADGLAALPLPAAP 2948
QY 948 --FRLEIDEPVLDHLVLRV--TERRAP--GLGEVEIAVDAAGLSFNDVOLALGMVPPD 1000
Db 2949 ALMRLEPGTDSLESLETLAARGAETLAPPLPGQYRIAIRATGLNFRDVLIALGMYPD- 3007
QY 1001 LRGKRPRLLLGECGRIVAVGEGVNGLVGCPVIALSAGAPATHVTTSALVLPBQA 1060
Db 3008 -----PALMGTEGAVTATGPGVTHLAPGDRVWGLLSGAVAPVADARTVAMRPG 3060
QY 1061 LSAIEAAMPVAVLTAMVALDRIARLQGERVLIHAATGAVGLAAVQMAQVGAETHATA 1120
Db 3061 WTFPAGASVPVFLTAIVVALRLDLADYKPERLLVHSAAGVGMBAVQLARHNGVHNGRA 3120
QY 1121 GTPKRAYLESIGV--RVVSDRSRDFVADVRAMTGGEGVDVVLNSGELLIDKSFNLIR 1178
Db 3121 SHGKMA--LRALGLDHAIASSRTLDFFESAFRAASGAGMDVVLNLSAREFVADLRLIG 3179
QY 1179 SHGRPELGRKRCY--ADNOLGLRPLRLNISFSLVDLRGMLEPRPAVRLLELLGLIA 1237
Db 3180 PGRFPEWKGKTVDRLDERVAADHPGVGYRAFDLGE-----AGPERIGMLLEVALFED 3233
QY 1238 GVFTPEPIATLPIARVADAFRSMAQHLGKLVLTG--DPEVORITPTAGA----- 1288
Db 3234 GYLRLPVTTWVRARDAFRHVSQARHGKVVLTWPSGLDEBGVLLTGGTALGGIYA 3293
QY 1289 ----- 1288
Db 3294 RHVGEWGVRLILVSRGTDAPAGELVHELEAGADVSAACDVADREALTAVLDSIP 3353
QY 1289 ----- 1288
Db 3354 AEHPILAVVHTAGVLSDGTLPSMTADVEHVLPRKVDAFLDELSTPGYDLAAFWMS 3413
QY 1289 ----- 1288
Db 3414 SAAAVFGAGGAGYAAANATLDAAMRRRTAGLPLSLGWMGLMAETSGMTGLSDTRSR 3473
QY 1289 ----- 1288
Db 3474 LARSGATPMDSETLTLDAAMRRDDPALVPIALDVAAALRAQORDGMLAPLLSGLTRGR 3533
QY 1289 -----GSTGDRLLDLRLAASAPAAALAEAFRTQSOVL--RTPEIKY 1332
Db 3534 VEGAPVNOGAAAGGGEADTDLGGRLAAMTPDDRAHLRDLVKTIVAVVLGHGTPS--RV 3592
QY 1333 GAELFTRLGMDSLMVELERNRIEASIKLSTTFSTSPNIALLAQNTLDALATA 1388
Db 3593 DIERAFRDGFDLSLTVELRNRLNAATGLRLPATLVFDPHTPELAGHLLDELATA 3648

```

RESULT 10
 T17409
 polyketide synthase type I - Streptomyces venezuelae
 C:Species: Streptomyces venezuelae
 C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #ext_change 17-Nov-2000
 C:Accession: T17409
 R:Xue,Y.; Zhao, L.; Liu, H.W.; Sherman, D.H.

```

Proc. Natl. Acad. Sci. U.S.A. 95, 12111-12116, 1998
A:Title: A gene cluster for macrolide antibiotic biosynthesis in streptomyces venezuelae
A:Reference number: 218773; MUID:9644533; PMID:9770448
A:Accession: T17409
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4613 <XUE>
A:Cross-references: EMBL:AF079138; NID:g3808326; PID:g380834; PIDN:AAC69329.1
C:Genes: pikA1
C:Superfamily: acyl carrier protein homology
C:Keywords: antibiotic biosynthesis; carrier protein
F:1010-1081/Domain: acyl carrier protein homology <ACP1>
F:2495-2566/Domain: acyl carrier protein homology <ACP2>
F:4407-4478/Domain: acyl carrier protein homology <ACP3>

Query Match 28.5%; Score 2051.5; DB 2; Length 4613;
Best Local Similarity 34.6%; Pred. No. 2.8e-102;
Matches 589; Conservative 183; Mismatches 515; Indels 417; Gaps 52;

QY 9 JAEDPIATVGAACRLPGGVLDLSGFWTLEGSRDVGVPAER-WDAAMFDPDDAPGX 67
Db 2601 ADDDPALVANSCTPIGDIRSPEDLMRLMSGEGITTPPTDKGMDLGLYDADALGR 2660
QY 68 TPVTRASFSDVACPDASFFGISPREALRMDPAHRLLEVCWALENAIAPSALVGTET 127
Db 2661 AVVREGFLHDAEERDAEPFGVSPREALMDPQGMILLTTSWEAVERAGIEPALSRSST 2720
QY 128 GVFTIGFSEYEAALPQATASAEIDAAGL--GTMPSGAGISIALGIRGCVAVDY 185
Db 2721 GVFTIGSYODVAAAPVAPRGVE---GYLLGSPSVASGRVATFGLEGATTVDTAC 2776
QY 186 SSSLVAVHLAOCSTLRSFGSTALAGVSLMSPSLWMLSKTRALARDGRCAFSABAD 245
Db 2777 SSSLTALHLAVALSGECTYALAGVAMMATPHMFVERSKRALAPDRSFAFSADAG 2836
QY 246 FGRGEGCAVVVLRKLSGARADDRILAVYRGSAINHDGASGLTVPNGSQEIVLRALA 305
Db 2837 FGAAGVGGLLVERISDARRNGHPVLAVVRGAVVQDGSNLTLPNGSQGRVIRQALA 2896
QY 306 DAGCAASVGVYEAHGTGTTGADPIEIQALNAVGLGSDVATPLLIGSYKTNLGHPEVAS 365
Db 2897 DARLAPGIDAVETGTGTSIDDPTEAOGLOATYKKEKPAERPLAIGSVKSNIGTQAAA 2956
QY 366 GTGLLKVLSLQHOGIPAHILHAQALNPRISGDLRLTYTRARTPMDNTRRARGVSF 425
Db 2957 GAAGITKVLAARHGTLPKTLHADPSPVADNSGLALVTPEIDMPAGTGRRAAVSF 3016
QY 426 GMSGTNAHVLEEAR-----AATCPAP-E 450
Db 3017 GISGTNAHVLEEARDAAGEVLEGADEVPEVSETVAMAGTSEVAEGSEAEAPAPGS 3076
QY 451 RPAEL-----LVLSARTASALDQAARLRDLHLETYPSOC-----LGDVAFSLATRS 497
Db 3077 REASLPGHLPVWLSAKEQSLRGQAALHAWISEPAALSDADGPARLRDVGTYLATSTR 3136
QY 498 AMEHRILAAVTSRREGRLAALDAAGOTSFGAVRSIADSSR--CKLAFLFTGGCAATLGMG 556
Db 3137 AFARHPAAVTAARDPDLGLATLQGGTS--AHVHLDTARGTTRAFLTGGSGSQPGAG 3193
QY 557 RGLYDWSAFREAPF-LCVRLFNOELDRPLREVMAEPASVDAALLDQTAFTQPLFTPE 615
Db 3194 RELYRHRPFAARALDEICHL--DGHLELPILDVMPAAGSAAALLDERRRYQCALPHE 3252
QY 616 YALAAALMSGVPEPLVAGHSTIGELVAAACVAGVSELEDAVFLVVAARGMLQALPRGAMV 675
Db 3253 VALFRLVSWGWRPALILGHSVGEITAAHVAGVFSLADAARLVAAARGRLMDELPGAMYL 3312
QY 676 STEAEADVAAAVAP---HAASVSIAAVANAPQOVVIAGAGQGVNHAIAAMARGARTKL 732
Db 3313 AVQAADETRVWLETREERAGRLDVAAVNGPAAALVSGADAAAEAAVWSGLGRTRRL 3372
QY 733 HVSHAHPFLMAPMLAEAGRVVAESVSYRRPSIVLVSNLSGKAC--TDEVSSPGYVWRHARE 791

```



```

Db      728 ORCEAEGRIARRIDVDYASHSAQVDAIREELIAALRGIEPRSTVAFFSTVGEIMDTAG 797
Qy      750 SSGYVWRHARREVFADGVKALHAAGAGTFVEVGPSTLGLVPAICMPD-----ARPA 833
Db      798 VNAEWTYSIRPQVPERAVRNAFDGQYRVFESSHPULIAGIEETLDCRGATGEI 857
Qy      834 LLAASRAGDEPATVLEALGGLMAYGCL-----VSMAGLPPS-GGRRVLPPTYPMORE 885
Db      858 VLP--LGRDD-----GGVGRFWLSAQAHVAGVDMRAAFADLGGRRVELPTAFANQ 910
Qy      886 RYVID----- 890
Db      911 REWLDGLGAVGDLGVGLVGAHEGLLAAVORPDGGVVLTGRI SVVAAPWLADHAYGP 970
Qy      891 -----TRADD----- 895
Db      971 VVLFETGVELALRAGDEVGCSVLQELTLQAPLVLPADGVRVQVVGVGSEOSTRANWV 1030
Qy      896 -----AARG----- 899
Db      1031 YSAGAGADSSPGWTLHAQGVLGVSQPAELSVMPVPGARAMDVAQVQLAARGYGYG 1090
Qy      900 ----- 899
Db      1091 PAFRGLOALMRGAEVADVTLPEGVPIRGFIPHAVLDALAHANGIVEGQOTMLPFSW 1150
Qy      900 -----DRAP-GAGHDEVE----- 912
Db      1151 QGVCHASGAARVVRRLAVGVGAVSVELADQGLPVSROLVMPVSAALSSSTAGD 1210
Qy      913 -----EG-----GAVRGCD----- 921
Db      1211 RGLLEMTVPVLEGGDIDGDAVWMLPRPHAGAQGDVLAAYVGVHEVLELQSWLAS 1270
Qy      922 -----RRAARLDHP----- 930
Db      1271 DATGLVVTTRGAVGVPDDVDTDLGAAVWGLVRSQAQEHPRVVLVDTGSAVEDAVG 1330
Qy      931 -----PRESGRREKVEAAGDRP-----FRLIDERGVL DHLVLTERR 969
Db      1331 FGARGGEFQVLVVRKGVTAARLAPVAAGLTLPASAGGRRLVAGGGTADVAVPVA 1390
Qy      970 APGLGEVEIADVAGLSFNDVOLALGMVBDLPKPNPPLLLGECAGRI VAVGCVGL 1029
Db      1391 ELATQOVRAVAVGVNFRDVLVALGMYPG--GE-----LGVGAGVVEVGCVTGL 1442
Qy      1030 VVGQPVIALSAGAFTHVTTSAALVLRPQALSALIEAAMPVAYLTAMVALDRIRLOPG 1089
Db      1443 AVGDVVMGL-LGLVGESEAVVDARLVTMVPAGMSLVEAAVPVAFLTA FYGLSLAEVAG 1501
Qy      1090 BRVLHAATGCVGLAVQMAQVGAHVNAHTACTPEKRAVLESLGR--VYSRSRDRFPA 1147
Db      1502 QGVVLHAGGVGVAALSLARWGAHVFTA-SRAKMDTLRAGMFDHDIHSRSRLEEBE 1560
Qy      1148 DVRAWTGEGVDVUNLSLSEGLIDKSFNLLRSHGFVELGKRD-----CYADNOGLRPF 1202
Db      1561 AFLRTBESSGVVVUNLSLAGETDASLRLLPBGGRFIELGKTDIDGQVVAERHNG---- 1616
Qy      1203 LKNLSFSLVDLRGMMLEBPAPVALLLELLGLIAAGVTPPPIATLPIARVADAFRSMQ 1262
Db      1617 VRYRAFVDIVE-----AGPDRIAMLSVGVGLLAGVLAARL PVKTFDARCAABA VRFVSQ 1670
Qy      1263 AOHGLKVLVTL----- 1273
Db      1671 AHHIKRVVLTLPDGGQSGSLAGTVVTGTGMAGSAVATHLVRRHGVANLVLSRGE 1730
Qy      1274 -----GDPEYQI-----RIP----- 1283
Db      1731 QADRAEVAALLRREGGAQVAVVSCVADRDALALLAGLDPRIYPLKGVFHAAGVLDAYI 1790
Qy      1284 ----- 1283

```

```

Db      1791 TGLPDRVDVTLRAKVDGAMNHLHETEDMDLSAFVFFSSMAGVGTPOAGVYAAANAFLD 1850
Qy      1284 -----THAGAPSTGDRDL--LD--- 1299
Db      1851 GLVAYRBRGLAGLSVANGMLEQASAMTRHLGERDRAMTQAGLAPLTTEQALGLDTRL 1910
Qy      1300 -----RIASAAPARAALAEFLR-----TQVSO---VLR---T 1327
Db      1911 QADRAVVAARLDRALLAGAALPALPFSQLAAGPTRRIRIDAADTAVMSGLVSRHALT 1970
Qy      1328 PE-----IKVGAELPFRLEMDLSMVELNRRIRIASLKLUS 1364
Db      1971 PERQRELTDVVISMAAVLGRSSVDINAKAFODLGFDSLTVAVELNRRRLKTAATGLTIS 2030
Qy      1365 TTFISTSPNIALAQNLLDALATALLSERRVAENLRACVQNDVFSSGADQ 1414
Db      2031 PTLIFDTPPTALAEHLDSRLVTA-----SSSDQ 2059

RESULT 12
A70984
probable polyketide synthase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jul-2000
C:Accession: A70984
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70984
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:molecule type: DNA
A:Residues: 1-2188 <COL>
A:Cross-references: GB:203857; GB:AL123456; NID:G3242252; PIDN:CAB06099.1; PID:G3242253
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: ppac
C:Superfamily: mycocerosic acid synthase, 3-oxoacyl-[acyl-carrier-protein] synthase I hc
nase homology; [acyl-carrier-protein] S-malonyltransferase; homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:56-460/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
F:1482-853/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F:1482-1770/Domain: long-chain alcohol dehydrogenase homology <SADH>
F:1802-1982/Domain: short-chain alcohol dehydrogenase homology <ACPI>
F:2072-2141/Domain: acyl carrier protein homology <ACPI>
F:2105/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 27.9%; Score 2014.5; DB 2; Length 2188;
Beet Local Similarity 31.7%; Pred. No. 1.1e-100;
Matches 565; Conservative 172; Mismatches 511; Indels 537; Gaps 35;

Qy      7 ERAAEDPIAIVGACRLEGGYIDLSGFPTLLEGRSDTGRVPAERMDAAAFDPDPDAPG 66
Db      30 EKSSSEPIAIVGACRLEGGYIDLSGFPTLLEGRSDTGRVPAERMDAAAFDPDPDAPG 89
Qy      67 KTVPTRASFLSDVAC--FDASFFGISPREALRMDPAHRLLEVCWALENAIAFSALVG 124
Db      90 TICSTEGFLLTSWQPDDEDAEFFSISPREAAMPQOGLLLEVALEADAGVPHTRIG 149
Qy      125 TETGVFIIGPSEYEALPQATASAEIDAAGGLGTMPVGVGGRISYALGAGPCVAVDTA 184
Db      150 TQTSFVGVTAYVDMVLTLAGRLRPVDLDAYIPTGNSAFAGRLAYITLGARPAVVIDTA 209
Qy      185 YSSSLVAVHLACQSLRSGECSTALAGVSLMISFSTVWLSTKTRALARDGRCKAFSABAD 244
Db      210 CSSSLVAVHLACQSLRSGESDMALVGTNLLLSGSPSIACGRKMWLSPEGCKTFDASAD 269
Qy      245 GFGRGEGCAVVVLRKLSGARADGDRILAVITGSAIINHDSAGSLTVPNGSGQETVLKRAL 304
Db      270 GYVRGEGAAVVVLRKLDADVDGDRILAVITGSAIINHDSAGSLTVPNGSGQETVLKRAL 329

```



```

Db 3336 HGRVRCGTC-----AGADIGFPAATASATSVLSGRVSYFFGLEPAITVDNAC 3383
QY 186 SSSLYAVHLACOSLRSCECSTALAGVSLMLSPSTLWMLKTRALARDRCFAFAEADG 245
Db 3384 SSSLYAVHLACOSLRSCECSTALAGVSLMLSPSTLWMLKTRALARDRCFAFAEADG 3443
QY 246 FERGGCAVVLKRLSGADADGRLIIVIRGSAIINHGAASSGLTVPNGSSQETVLKRALA 305
Db 3444 TMAEVEGVLIVERLSDAQAKHQVLAIVRSSAVNDDGASNGLTAPNGSQQRVIOALIS 3503
QY 306 DAGCAASSGVYEAHGTCTTGDPIELIOALNAVYGLGRVATPRLIGSKTKLGHFEVAS 365
Db 3504 NGGLAAHEVDVVEAHGTCTTGDPIELIOALNAVYGLGRVATPRLIGSKTKLGHFEVAS 3561
QY 366 GTTGLKLVLSLOHGOIPALHQAOLNPRIISWCDLTLVTARATPMPDNTPRAGVSSF 425
Db 3562 GVSIGVTKMMALOHTVPTLHVDEPSRIVDMAAGAVELRENOQPIEGTDRPRRACVSSF 3621
QY 426 GMSGTNAHVLEBPATCTPPAPE--RPAEL-----LVLASATASALDAQAARLFD 475
Db 3622 GVSIGTAAHVILIESAPPA---QPAEBAQPVETTPVAVSDVLPVLSAKTQPALTEHEDRLA 3678
QY 476 HETVFSQCLGDVAVSLATTSGMEHRLAVATSRGELRALDAAAGCQ--TSPCAVRSI 533
Db 3679 YLAASPGADIRAVASTLATVRSVFHRAVLG-----DDTVTGAIVTDP----- 3722
QY 534 ADSRKGKLAFLFTGCGAQTLLGMRGLYDVMSAFREAFDLCVRLFNQELDPLREVMMAEP 593
Db 3723 -----RIYFVFPQGQWMLGMSALRDSVVFPAEMACAAALREFVWMDLFTVL----- 3772
QY 594 ASVDAALLDQTAFTQPALFTFEYALALWRSWGEVELVAGHSIGELVAVCAVAFSLSD 653
Db 3773 --DPAVVDVAVVQPSWAMWVSLAAVQAGVREDAVIGHSQGIILAAVCAVAFSLSD 3830
QY 654 AVFLVAARER-LMQLPACGAMVSTAEPEADVAANAAPAAVSTIAVAPDOVYIAGG 712
Db 3831 AARIYTLRSQALIRAGIAGGAAVSLPAQDVEL-----VDGAMIILAHNGPASTYIAGRP 3885
QY 713 QPVHAIILAAAMARGARTKALHVSHPSPMLMAPLEAFGRVAESVYRSPSTVLVNSLNG 772
Db 3886 EAVDVHVLTHHEAGVAVRRIYDYASHTHVELIRDELIDITSOSSQTPVLPWISTVVG 3945
QY 773 KACTDEVSSPGYVVRHAREVVRPADGVKALHAAAGCTFEVEVGPKSTLLGLVPACMEDARP 832
Db 3946 -TWVSPDLGEXWYRNLREPVRGHPAVSOLQAGDVFVEVASPVLLQAM-----DDVV 3999
QY 833 ALLASSRAGRDEPATVLEALGLMAVGLVSMAGLFPSCGRVY-PLPTYPKQREKRYWIT 891
Db 4000 VTVATLRRDDGDATRMILTALQAUVHGVTVDPWPAITGTTTTLVLDLPTVAFQHQRYWLS 4059
QY 892 -----KADDAARGRRAPGAGH----- 908
Db 4060 VDRAAADGHPILGTVVALLPESGCVLITGVSLATHTWMLADHVRGSLVLLPGTAFVELVVR 4119
QY 909 --DEVE----- 912
Db 4120 AADEVCEVDVDELVIETPLLLPQTGGVQLSVSAEADSGHRTVTVFSQADNTDAIRHV 4179
QY 913 -----EG-----GAVRGCD-- 921
Db 4180 SATISTSDTPLSPESDTPWPAQAQPMNVADFYDRLMAAGYEXGAPQGLQAMRGGDV 4239
QY 922 -----RSARLD--HPP-----PSSGRKREVEAAGD-- 945
Db 4240 YAEVVALAEQAQGAARYDVHAPLLDAALHACTLNTPDAEBGVGLPFSNMHVQVHAATPAT 4299
QY 946 ----- 945
Db 4300 LRAVAVTQEGADGWSLRVADDTGRVATIGSLITRPVATDLTGTTDDLTLTWTVIIPPOE 4359
QY 946 -----RPR 949
Db 4360 TSLPMVRYEDLADGNMPVDPVVVFTALPNSSENPLGPRAPGADPLAOPDTDROVLKPPQ 4419

```

```

QY 950 -----LE 951
Db 4420 TWLPCTFTDSTLVATGTGGLAAAVSGLMRGAQSHRGRFVLESDDTLTLRQLAATVG 4479
QY 952 IDEP-----GVLDHLVLRVTERRA 970
Db 4480 LDEPRLINDGREYEAARLARITTTPTEDPTPLTIPDSRAWLLEOPRGSTLODILLVPTDPAE 4539
QY 971 PEL--GEVEIIVDAGLSFNDVQALAGVPPDLPEKPRPPLLGGECAGRIYAVEGVNG 1028
Db 4540 QPLRGEVRIDVRAAGLFRDVLVAGVNDRP-----TGSEAGVLEVEGPDVD 4591
QY 1029 LVVQGVIALSAGAPATVTTSAALVLRPQALSAIEAAMPVAVLITWYALDRITARIOP 1088
Db 4592 LARGDRVGVGSGSGPRAADRRLAVIIPDQWSTTAAVSVVATVATVYGLVDLAGLSE 4651
QY 1089 GERVLIIHAATGGVGLAAVQMAOHVGAEVHATAGPEKRAYLES-LGVRYVSDRSRDPVA 1147
Db 4652 GESVLIIHAAGGVGAATQIARHLGARIVATSTGKHILREAGLEDTIADSRITLSPQE 4711
QY 1148 DYRANTGEGVDVUNLSISGELIDKSPILASHGRFVELGRDCTADNQLG--REPLRL 1206
Db 4712 TFLANTHOGVDVUNLSISGDFVDASLDLPRGGRFVEMGKTIDPPOVADIRFGTTYQ 4771
QY 1207 SFLVDLRGMLEBPARYALLLEELGLIAGVFTPPRIATLPRIARVADAFRSMQAQHL 1266
Db 4772 AFDLMD-----AGDRILREITELLTLFTQGVLLPLVQAMDIDQARDAFSMBGRARI 4825
QY 1267 GKVLITL--GDPEVOIRIPTHAGAFSTGDRDLIDRLASAPARAALAEFLTFVSO 1323
Db 4826 GKIVITLPRPPDDITLIT--GSGSVLAG--ILARHLLAEHGARHLL--LSRTTPQ 4878
QY 1324 VLRTEIVGAALFTRGMSMAVELNRLEASLXLKSTFTSTPNIALLAQNLLD 1383
Db 4879 ALIKELAEIAGH-----VDATC-----DVSDRGLARVLAVGSPHEPLTA----- 4919
QY 1384 ALATALSLEVAENLRAGVQNDPVSQADQDMEI 1418
Db 4920 VHTHAGALDDGVESLTTQODTVLRPKADGMHL 4954

```

```

RESULT 14.
T17463
t1famycin polyketide synthase modules 1-3 - Amycolatopsis mediterranei
C1Species: Amycolatopsis mediterranei
C1Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C1Accession: T17463
R1Schupp, T.
submitted to the EMBL Data Library, December 1997
A1Reference number: Z18802
A1Accession: T17463
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-4735 <SCH>
A1Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227120; PIDN:CA11035.1
C1Experimental source: strain LBG A3136
C1Superfamily: acyl carrier protein homology; acetate-CoA ligase homology
C1Keywords: carrier protein
F153-500/Domain: acetate-CoA ligase homology <ACP>
F1543-610/Domain: acyl carrier protein homology <ACP1>
F1202-2173/Domain: acyl carrier protein homology <ACP2>
F13079-1150/Domain: acyl carrier protein homology <ACP3>
F14578-4649/Domain: acyl carrier protein homology <ACP4>

```

```

Query Match 27.8%; Score 2002.5; DB 2; Length 4735;
Best Local Similarity 37.9%; Pred. No. 1.3e-99;
Matches 586; Conservative 183; Mismatch 526; Indels 251; Gaps 50;
QY 2 ADRIE---RAEDRIAVGASCRIPGVYIDLSGWITLBSGRDVGAVPAR-WDAAM 57
Db 2182 ANRPAEIGTAAEEPIAVAACRPFGVHSPEDLMRLVADGADAVTEFPADRGWDTDR 2241

```

QY 58 FDPPDPAPGKTPYTRASFILSDVACFDASFPGISPREALRMDPAHRLILEVCWEALENAAI 117
 Db 2242 YHDDPHEGTTTYRHAFGLDDAAGFDAAFPGISFNEMLANDPOORLILETSMELFERAAI 2301
 QY 118 APSALVGTETGTFIGIPSEYEALPOATASAEIDAHGIGTIPSYGAGISTALGRP 177
 Db 2302 DPTTLGADGIGFAGVAGSHSYMRMHPA---AGVEGFRITGCGASVLSGRVAHVFGECP 2358
 QY 178 CVAVDPYASSSLVAVHAAQSLRSGECSTALAGVSLMSPTLVMLSKTRALARDGRCK 237
 Db 2359 AVAVDPYACSSSLVAHAAVQALRGCSMALAGVVMGTFETFEFERSRORGLAPDRCK 2418
 QY 238 AFSAEADGFGRCGCAVAVVLRKLSGARADGRILAVIRGSAIHHDGASGLTVPNGSSOR 297
 Db 2419 AFADGADGTGWSGVGLLVERLSEAGORGHQVLAVRGSAAVSDGASNGLTAPNGSSOQ 2478
 QY 298 IYLRALADAGCAASSVGVYEAHGTGTTGDPTEIOALNAVYGLGRDVPPLIGSYKTN 357
 Db 2479 RVIRKALAAAGLSTSDVAEAGTGTTLGDPTEAEALLATYCGNRE--TPMLIGSYKSN 2536
 QY 358 LGHPEYASGITGLKVLSTOHQOIPAHHAQALNPISTW--GDRLTIVTARTPMPDMN 415
 Db 2537 LGHTQAAAGVAGVAKVMAMRHGVLPRTLVDRPSYVDWSAGAVEL--LTGAR--DWVSG 2594
 QY 416 TPRRAGVSSFGMSGTNAHVLEEAAPATCTPPAPERPAELIV--LSARTASALDAQAR 472
 Db 2595 HPRRAGVSSFGIGTNAHVLEEV--AAPITPOPE--PAEFLVPLVSAKRIAPAGRGAGR 2652
 QY 473 LRDLLETYPSCQGDVAFSIATTSRAMEHRLAVALTSREGURLAALDAAOQTSFGAVRS 532
 Db 2653 LAFLGRTDVRVDAAYALATTRAQDLHRAVVLASRAOLCADLAAGSGVATGTPV-- 2710
 QY 533 IAASSRCKAFLFTGCGAOTLGMGRGLYDWSAFREAFDLCVLFNOL--DRPLREYMA 591
 Db 2711 ---DKKALVLTGQSSQWAGRELAETFPVFRDFAEACBAVDTHLERPREVED 2765
 QY 592 EPASVADALIDQTAFTQPALFTFEYALAAALMRSGVPELVAGHSIGELVAACVAGFSL 651
 Db 2766 ---DSALLDQMTYOGALFAVETALFRLFESGAVRGLLAGHSIGELAAAHVSGVLDL 2820
 QY 652 EDVFLVAARGLMOALPAGAMVSIAPADVAALAAVAPHAASVSIANVAPDOVITAGA 711
 Db 2821 ADAGELVAARGLMOALPAGAMVAVOATEDEVAFLP---DGTVCVAANVGPPSVVLISGT 2877
 QY 712 GQVHAIAAAMAARGARTKALVSHAFSPMLAMPLEAFGRVAVESVYRRPSITVNLTS 771
 Db 2878 EAALVAADLADLAGRKRTRRLAVSHAFSPMLMEPLDFAVAVERLTYRAGSLFVNSTLT 2937
 QY 772 GKACTDEVSSPGVVRHAREVVRPADGVKALHAAGAGTFVEVGKSTLGLVPAKMPDAR 831
 Db 2938 GELAA--LDSPDVYVQGVNANVRSDAVTLGAGASTFLEIGCGALAAAMAGTIGGPE 2995
 QY 832 PALLASSRAGRDEPATVLEALGIMAVGVNAGLF--PSGGRVPLPTYPMQREYWI 889
 Db 2996 QSCVATIRKKGAEVDPVLTALAEHLVAGVGVDTVTVADEPATVAGTVPYAFQHQRFVW 3055
 QY 890 DTRADDARQDRAPAGH-----DEVEGCAVRRG--DRSARLDHPRESG-- 935
 Db 3056 D--VDTEAAVSVTPPPAPPIVDRPVQVLELVRESAAVVLGHRPAGFDDRSFKDGF 3113
 QY 936 ---RREKVENAG--DRPFRLEIDEPG--VLDHLVLRTERRAAGLGEVEITAVNAG 984
 Db 3114 SLASVAKRNLRLDQTVGLPSTLTFDYPNPAVLADHL-----RAELIGERAPAPRV 3166
 QY 985 LSFNDVOLALGMPDDLPGKPNP-----LLLGSECA--GRIVAGSGVGLVGOPIA 1037
 Db 3167 RDVSDDEPIATVIGMSTRLPGGDSPEELMKLVAEGRDVSGPVRGMDLGLYHPDRAHA 3226
 QY 1038 LSAQAFTHVTYSALVLRPQ-----ALSATFAAM--PAVYL---TAWALDRIRALQ 1087
 Db 3227 ---GTSYTRSGGFLHDAAGFAGLISPREALAMDPOORLILETSMELER--AGVD 3279
 QY 1088 P---GBRV---LIH-----AAT 1098

Db 3280 PLASGSDVGFETGIYHNDVYTRLEVPEDVOGYTMTGTASSVAGRAVYVEGEPAPT 3339
 QY 1099 GGVGLAAVQACHVGAEV-----HATAG-----TPRKAYLSIGVRYVS--DSRSOR 1144
 Db 3340 VDTACSSLVAAHMLAAQALRGCSMALAGATVNASPD--AFLESRORGLSADORCA 3397
 QY 1145 FV--ADVRAMTGEGVDV-----VUNSLSGELIDK---SFNLRSHGRFVEL 1186
 Db 3398 YAEAGDGTGMAGVGIVVLELRSVARERGRVLAIVLRGSAVNOGASNGLTAPNGSSOQ 3457
 QY 1187 GKRDCTADNQLGRPLRLNLSFLYDL-----RGMLEPAPVALL-----BEELGL 1234
 Db 3458 VIRGALAA--SAGLAP-----SDVDVVEGHGTGALGDPVQALLATYGOEREPML 3508
 QY 1235 -----TAAGFTTPPIATLPIARVADAFRS--WAQOHIGKVLTLGDPEVOIR 1281
 Db 3509 GSIKSNLGHTOAAAGV-----GVYIMAMRHGVNPAVLHDERISQVDSAGALE 3560
 QY 1282 IPTHAGAPSTG-----DRDLRLASNAAP--AARAAA 1312
 Db 3561 VLTEAREWPRTRGPRRAGVSSFGASGTNAHLIEBGPAAEAVDEEAVSVPLVARSAG 3620
 QY 1313 LEAFRLTQVSOVLRTPEIKVGALEFT--RLGMDSLMAVELRNRIRIA 1357
 Db 3621 SLAQAGRLAAVLENSLAGVAGALVSGRATINERAVIAGSRDEA 3666

RESULT 15
 T30225
 polyketide synthase - Streptomyces hygroscopicus
 C:Species: Streptomyces hygroscopicus
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #extl_change 17-Nov-2000
 C/Accession: T30225
 A:Particic, J.F.; Molnar, I.; Schwecke, T.; Koenig, A.; Haydock, S.F.; Kraw, L.E.; Strau
 Gene 169, 9-16, 1996
 A:Title: Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hy
 A:Reference number: Z20782; MUID:96186896; PMID:8635756
 A/Accession: T30225
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-10223 <ADP>
 A/Cross-references: EMBL:X86780; NID:9987088; PID:9987099; PIDN:CAA60459.1
 C/Genetic:
 A:Gene: rapB
 C/Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acyl carrier protei
 C/Keywords: carrier protein
 F:54-449/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>
 F:1533-1604/Domain: acyl carrier protein homology <ACP1>
 F:1647-2042/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>
 F:2135-2409/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT1>
 F:3172-3243/Domain: acyl carrier protein homology <ACP2>
 F:3287-3661/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F:3718-4052/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT2>
 F:5160-5231/Domain: acyl carrier protein homology <ACP3>
 F:5275-5667/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS4>
 F:6176-6847/Domain: acyl carrier protein homology <ACP4>
 F:6891-7255/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS5>
 F:8411-8442/Domain: acyl carrier protein homology <ACP5>
 F:8526-8921/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS6>
 F:9012-9285/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT3>
 F:10069-10140/Domain: acyl carrier protein homology <ACP>

Query Match 27.7%; Score 2000.5; DB 2; Length 10223;
 Best Local Similarity 30.5%; Pred. No. 4,6e-99; Indels 521; Gaps 35;
 Matches 562; Conservative 197; Mismatches 562;

QY 5 PIRRAED--PIATVAGSCRPGGVNIDLSGFWTLLGSSRPVGRVPAER--WDAAWFDPDP 62
 Db 3258 PVRVAVQDEPLATVGAACRLPGGVSSPBDLMRLVEGTDAVSGFPIDRGNDVGLRDPDR 3317
 QY 63 DAPGKTPVTRASFILSDVACFDASFPGISPREALRMDPAHRLILEVCWEALENAAIAPSA 122

```

Db 3318 DAAGSYAEGGELDTAAGPDAGPFGISPREALANDPOORLILEVSWENEFERAGIEPGSV 3377
Qy 123 VGTETGVIGIGPSEYEALPOATASAEIDAHGGLG--TMPVSGAGRISTYALGRLGPCVA 180
Db 3378 RGSDDGVITGAFPVGYGAGA-----AREGIGATPAAPVNLGRLSTYFGLGCPAIT 3427
Qy 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMSPTLWMLKTRALARDGCKAFS 240
Db 3428 MDTACSSSLVHLALHQAOLRNGECSMALAGVTVATPEVFTEFARQGLADGCKKAPA 3487
Qy 241 AEADEFGEGCAVVVVKRLSGARADGRIILAVTIGSALNHDGSSGLTPVNGSQEIVL 300
Db 3488 DSADAGSEEGGLLVERLSLSPARRNGHQLAVVGSANQDASNGFTAPNGPQOQRYI 3547
Qy 301 KRALADACCAASSVGVVEAHGTGTLGDPRIEIQALNAVYGLGRDVAITPLIGSVTNIGH 360
Db 3548 ROLANAGLTTRVEVVEAHGTGTLGDPRIEQAQIATYGORE--QPLIGLTLSSNGH 3605
Qy 361 PEYASGITGLKVLVSLQHQOI PAHLHAQALNPRI SWGDLRLTVTRARTPMDMNTPRRA 420
Db 3606 TOPAAGVSGVIMVVALQHSITVPRTLHVNEPBRHVDMSAGAVELVTENQSWVTGRPRRA 3665
Qy 421 GVSAGSGMGTNAHVLEAPAA---TCPPA-PRPRAEL--LVISARTASALDQAARL 473
Db 3666 GVSAGVSGTNAHVLEAPAAQSVNNAPVATPVVASELVELVISAKTLPALTEHEDRL 3725
Qy 474 RDHETYSQCLGDVAFSLATTRSAMEHRLAVALTSREGLRALDAAAGQGPSGAVRSI 533
Db 3726 RAYLAASGADMRVAVGSTLALTRSVFERH-----AVLLGHDTVTVTG 3769
Qy 534 ADSRGLKLFETGGAQTLGNGRGLYDWSAFREAFDLCVRLFNOELDRPREVMAAP 593
Db 3770 TAVSNPRVVFVPGQGMWLGMSALRGSVVFARMAECAALAEFMDWDFATL----- 3825
Qy 554 ASVDALLDQRTFOTPALFTEYIALAALMRSGVPELVAGHSIGELVAAACVAGFSLD 653
Db 3826 --DPAVVDVAVVQPASMAVWVSLAAVWQADGVPRDVTIGHOSEIIAACVAGAVSRD 3883
Qy 654 AVFLVAAGR--LMQALPAGGAMVSI EADVAAVAAPHAASVSI AAVNAPDOVVIAGG 712
Db 3884 AARSTTLRSQAIARGLAGGAAASVALPAHEIEL-----VDGAWIAAHNGPSTVAGAP 3938
Qy 713 QPVHAIAAAMARGARTALVHSHA FHSPLMAPLEARGVAVESVYRPSIVLSNLG 772
Db 3939 EAVDRVLVHEARGVRVRIANDVASHTRPVELIDELLDITAGIGSQAPVPWISTYDG 3998
Qy 773 KACTDEVSPP--GTVWPHAREVVRPADGVKALHAAGCTFVEYEPKSTLGLVPACMD 829
Db 3999 ---TWVEGPLDVEYVYRNLREPVGDSAVQOLRAEGDVFVEVGSASPVLLQAM-----D 4049
Qy 830 ARPALLASSRAGRDEPATVLEALGGLMANVGLVSWAGLFPSGGRVP--LPTVPMQREXY 888
Db 4050 DDVVTVAITLRDDGATRLTLALQAFVEGVTVDPALIGTITTRVPDLPTAFQHORFW 4109
Qy 889 IDTK-----ADDAARGDRAPAGH-- 908
Db 4110 AEGADRSVAGHPLGLVAVELPESDGVLLTGRVSLATHAMLDHVRGSVLLPRTGFVEL 4169
Qy 909 ---DEV----- 911
Db 4170 VVRADEVCVDELVIETPLLPSSASVHLVSVGEADESGRGVTVFSRADGADAMT 4229
Qy 912 ---BEGAVRG----- 919
Db 4230 RHVSATIGVGAALSLPELAMPPAQAOPVGLGDFYDRLTGAGYEGYGAFOGLQAMRDG 4289
Qy 920 --- 919
Db 4290 DTVFAEVALAEQAEARFAVHPALDLDAALDAGITLNTLDNAEQVRLPFSWNGVQVART 4349
Qy 920 --- 919
Db 4350 GSATLRVADNPTSDGMSVRAVDSDGRPVATVDSLVTBPVATDLGSAADLLTJVWTEIP 4409

```

```

Qy 920 ----- 919
Db 4410 TPQOTGLSVGRFEDLDADGVPEVNVCTALPDSESNPLALDPDPLVQTRITLTQVLO 4469
Qy 920 -----GDR-----RSARLDHP----- 930
Db 4470 AVQAWLAGERFTDSTLVRTGTGLATAGVSGLMRSAGSEHGRFVLVECDNLTLOQAA 4529
Qy 931 -----PPSGRREKVPA-----GDPFFLEIDEPEVLDHVL--RVT 966
Db 4530 TVGLDEPRLVCDGRFEVPRLARANTPSSSLPTTIPGDAMLLBQSHSGTLRLDALVPAET 4589
Qy 967 ERRAPGLGEVEIAVDAAGLSFNDVOLAGWPDDLPGRPNPLLIGKGCAGRIVAVGCV 1026
Db 4590 AERPLQSEVAVDVRAGLNRFDVILALGTYPGF-----AVIGABAAGVLEVEGV 4641
Qy 1027 NGLVVQOPVIALSAGAFTHVTTSAALVPRPQALSAIEAAMPVAYLTAMVALDRIARL 1086
Db 4642 QDLAPGRDVFGLVGGFGCAVALIADRMGLVLPDGMSPFTTAAVVPVFPATAYGLVLAGL 4701
Qy 1087 QPGERVLIHAATGVGLAAVQMAQVGAENVATAGTPKRAVLES--LGVRVSDSRDRF 1145
Db 4702 SAGESVLIHAAAGGVMAATQIARHLGARIVATASTGQHVLRXGLDARIIGSRTTG 4761
Qy 1146 VADVRAWTGEGGVVNLNSLSELIDKSPNLLRSHGRFVELGKRDVADNOLGL--RPLR 1204
Db 4782 REMULDTDSRGVDVNLNSLSDGFVDASLDLLPRGGRFVENKTDIRPHOVYADRPCTS 4821
Qy 1205 NLSFSLVDLRGMLERPARVALLBEILGLIAAGVFPPPIATVPIARVADAFRSMAQAO 1264
Db 4822 YQAFDLM-----AGPRLREITIALDLALFAQGLVLPVRAMDIRQAREAFSWSRAR 4875
Qy 1265 HLGKLVLTGDPVEQIRIPTAGAGPS---TGDRDLRLASAPAAALAEALFRTQV 1321
Db 4876 HIGKIVLT-----VPOQLDANGVTLVYGGSGVL-----AGIAARHLVAEOGVRL 4921
Qy 1322 SOVLRTEPKVGAELFTRLGMSLIMAVELNRIEASL-----KLKSTTFSTSPNAL 1376
Db 4922 LLSRSTPD-----DALINELG-----ELGARVDTALCVSDRAGLARITAGVSEHPL 4969
Qy 1377 LAQNLDALATLSLERVAENLRAGVONDEVSSGADQDWEI 1418
Db 4970 TA-----VIHTAGALDDGVESLTAQOLETVLRPRADGAWHL 5006

```

Search completed: October 2, 2003, 17:35:08
 Job time : 93.4261 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 2, 2003, 17:34:03 ; Search time 81.0897 Seconds
(without alignments)
2772.494 Million cell updates/sec

Title: US-10-014-717-2
Perfect score: 7210
Sequence: 1 VADRPPIERRAEDPIAIVGAS.....GVQNDPVSGADPDWEIHAL 1421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7210	100.0	1421	14	US-10-014-717-2
2	4137	57.4	2439	14	US-10-014-717-7
3	3311.5	45.9	7257	14	US-10-014-717-5
4	2311.5	32.1	6146	15	US-10-156-761-10436
5	2249	31.2	6145	15	US-10-156-761-7963
6	2245.5	31.1	3798	14	US-10-014-717-6
7	2199.5	30.5	3352	14	US-10-156-761-7961
8	2197.5	30.5	3564	15	US-10-156-761-7964
9	2159.5	30.0	1746	15	US-10-156-761-7963
10	2152	29.8	1835	15	US-10-156-761-7963
11	2149.5	29.8	1784	15	US-10-205-032-18
12	2128	29.5	3816	11	US-09-808-880-3
13	2125	29.5	5215	10	US-09-861-289-2
14	2125	29.5	5215	10	US-09-860-846-2
15	2125	29.5	5215	11	US-09-988-384B-2

16	2125	29.5	5215	11	US-09-836-821-2	Sequence 2, App11
17	2114.5	29.3	3745	15	US-10-205-032-14	Sequence 14, App1
18	2105	29.2	3939	15	US-10-156-761-10434	Sequence 10434, A
19	2077.5	28.8	3613	15	US-10-156-761-10432	Sequence 10432, A
20	2052.5	28.5	3739	10	US-09-861-289-33	Sequence 33, App1
21	2052.5	28.5	3739	10	US-09-860-846-33	Sequence 33, App1
22	2052.5	28.5	3739	11	US-09-988-384B-33	Sequence 33, App1
23	2052.5	28.5	3739	11	US-09-836-821-33	Sequence 33, App1
24	2052.5	28.5	11877	10	US-09-861-289-6	Sequence 6, App11
25	2052.5	28.5	11877	10	US-09-860-846-6	Sequence 6, App11
26	2052.5	28.5	11877	11	US-09-836-821-6	Sequence 6, App11
27	2052.5	28.5	12199	11	US-09-988-384B-6	Sequence 6, App11
28	2051.5	28.5	4551	11	US-09-793-708-1	Sequence 1, App11
29	2051.5	28.5	4551	12	US-10-201-365-2	Sequence 2, App11
30	2051.5	28.5	4551	12	US-10-160-539-1	Sequence 1, App11
31	2051.5	28.5	4613	10	US-09-861-289-31	Sequence 31, App1
32	2051.5	28.5	4613	10	US-09-860-846-31	Sequence 31, App1
33	2051.5	28.5	4613	11	US-09-988-384B-31	Sequence 31, App1
34	2051.5	28.5	4613	11	US-09-836-821-31	Sequence 31, App1
35	2051.5	28.5	4840	15	US-10-156-761-10435	Sequence 10435, A
36	2048.5	28.4	3739	11	US-09-793-708-2	Sequence 2, App11
37	2048.5	28.4	3739	12	US-10-201-365-3	Sequence 2, App11
38	2048.5	28.4	3739	12	US-10-160-539-2	Sequence 2, App11
39	2046.5	28.4	4881	15	US-10-156-761-8481	Sequence 8481, Ap
40	2025	28.1	1925	15	US-10-205-032-12	Sequence 12, App1
41	2024.5	28.1	3970	15	US-10-156-761-10429	Sequence 10429, A
42	1966.5	27.3	6396	12	US-09-940-316B-72	Sequence 72, App1
43	1954.5	27.1	1832	14	US-10-014-717-4	Sequence 4, App11
44	1918.5	26.6	4150	11	US-09-808-880-2	Sequence 2, App11
45	1889	26.2	3519	11	US-09-808-880-4	Sequence 4, App11

ALIGNMENTS

RESULT 1
US-10-014-717-2
Sequence 2, Application US/10014717
Publication No. US20020192778A1
GENERAL INFORMATION:
APPLICANT: Schnupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Rose
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OR INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/10/014,717
CURRENT FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: US/09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1421
ORGANISM: Sorangium cellulosum
US-10-014-717-2
Query Match 100.0%; Score 7210; DB 14; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VADRPPIERRAEDPIAIVGASCRIPGVYIDLSGFWTLLEGSRDIVGRVPAERMAAANFDP 60
DB 1 VADRPPIERRAEDPIAIVGASCRIPGVYIDLSGFWTLLEGSRDIVGRVPAERMAAANFDP 60
QY 61 DPDAFGKTPYTRASFSDVACPDSPFGISPREALRMDPAHRLILVYCWELNAATAAPS 120
DB 61 DPDAFGKTPYTRASFSDVACPDSPFGISPREALRMDPAHRLILVYCWELNAATAAPS 120
QY 121 ALVETGVETIGIPSEYEALPQATASAEIDAHGIGLTPSVGAGRISYALGLRGPVA 180

Db	121	ALVETGETGFIIGISEYEPALLPQATASABIDHAGGLGTPSPGAGISTALGRCVA	180
Qy	181	VDTASSSLVAVHLACQSLRSRSECTALAGVSLMSPSTLWVLSKTRALARDGRCAFS	240
Db	181	VDTASSSLVAVHLACQSLRSRSECTALAGVSLMSPSTLWVLSKTRALARGRCAFS	240
Qy	241	AEADPFGREGCAVVVVKRLSGARADGDRILAVIRGSAINHDAAGSLTVPNGSQEIVL	3000
Db	241	AEADPFGREGCAVVVVKRLSGARADGDRILAVIRGSAINHDAAGSLTVPNGSQEIVL	3000
Qy	301	KRALDAGCAASSVGVVEAHGTTGLGDPLEIOLNAVYVGLGDVATPLLIGSVKTMIGH	3600
Db	301	KRALDAGCAASSVGVVEAHGTTGLGDPLEIOLNAVYVGLGDVATPLLIGSVKTMIGH	3600
Qy	361	PEYASGITGLKVYLSLQHQI PAHLHAOLNPRISMGDLRLVTBARTPWPMMNTPRA	4200
Db	361	PEYASGITGLKVYLSLQHQI PAHLHAOLNPRISMGDLRLVTBARTPWPMMNTPRA	4200
Qy	421	GVSSFGMSGTNAHVYLEAPATCTEPAPERPAELLVLSARTASALDAQARLDHLETY	4800
Db	421	GVSSFGMSGTNAHVYLEAPATCTEPAPERPAELLVLSARTASALDAQARLDHLETY	4800
Qy	481	PSQCGDVAFSLATTRSAEMERLAAVATSBEGIRALDDAAOQOTSPGAVRSIADSRCK	5400
Db	481	PSQCGDVAFSLATTRSAEMERLAAVATSBEGIRALDDAAOQOTSPGAVRSIADSRCK	5400
Qy	541	LAFLFTGGOAQTGLMGWRGLDYVWSAFREAFDLCVRLFNELDBPLSEVMAEPAASYDAL	6000
Db	541	LAFLFTGGOAQTGLMGWRGLDYVWSAFREAFDLCVRLFNELDBPLSEVMAEPAASYDAL	6000
Qy	601	LDQTAFTOPALTFEYALALWRSWGBEDELVAHGSIGELVAA CVAGVSLDVAFLVAA	6600
Db	601	LDQTAFTOPALTFEYALALWRSWGBEDELVAHGSIGELVAA CVAGVSLDVAFLVAA	6600
Qy	661	RGRMLQALPAGAMWSIEAPADVAAA VAPHAASVSLA VNPDOVVIAGAGOPHATA	7200
Db	661	RGRMLQALPAGAMWSIEAPADVAAA VAPHAASVSLA VNPDOVVIAGAGOPHATA	7200
Qy	721	AMARGARTKALHVSHPFSLPMLAEFGRVAESVSRPSITVLSNLSGACTDEVS	7800
Db	721	AMARGARTKALHVSHPFSLPMLAEFGRVAESVSRPSITVLSNLSGACTDEVS	7800
Qy	781	SPGYVWRHAREVVRPADGVKALHAAGCTFEVEGPKSTLLGLVPACMPDAPRALSSRA	8400
Db	781	SPGYVWRHAREVVRPADGVKALHAAGCTFEVEGPKSTLLGLVPACMPDAPRALSSRA	8400
Qy	841	GRDEPATVLEALGILMAVVGGLYSMAGLFPSSGGRVLPPTYPMQOREXVIDTKADDAKGD	9000
Db	841	GRDEPATVLEALGILMAVVGGLYSMAGLFPSSGGRVLPPTYPMQOREXVIDTKADDAKGD	9000
Qy	901	RRAPAGHDEVEEGGAVRGDRRSARLDHPPPSGGRKEXVLAAGDPFRPLEIDEPGLDH	9600
Db	901	RRAPAGHDEVEEGGAVRGDRRSARLDHPPPSGGRKEXVLAAGDPFRPLEIDEPGLDH	9600
Qy	961	LVLRVTERRAPGLGEVEILAVDDAAGLSFNVOVLALGMVPPDLPGKPNPPLLIGSECAGRIV	10200
Db	961	LVLRVTERRAPGLGEVEILAVDDAAGLSFNVOVLALGMVPPDLPGKPNPPLLIGSECAGRIV	10200
Qy	1021	AVGEGVNGLVGQPVITALSAGAPATHVTTSSAALVLEPRQALSAIEAAMPVAYLTAWYAL	10800
Db	1021	AVGEGVNGLVGQPVITALSAGAPATHVTTSSAALVLEPRQALSAIEAAMPVAYLTAWYAL	10800
Qy	1081	DRIAPLOGERLTHAATGCVGLAAVQMOHQHGAETHAATGPEKRAYLESIGVRYVSDS	11400
Db	1081	DRIAPLOGERLTHAATGCVGLAAVQMOHQHGAETHAATGPEKRAYLESIGVRYVSDS	11400
Qy	1141	RSDFEVADVRAVATGEGVNVVNLNSLSELIDKSFNLLRSHGREVLEKSDCYADNOLGLR	12000
Db	1141	RSDFEVADVRAVATGEGVNVVNLNSLSELIDKSFNLLRSHGREVLEKSDCYADNOLGLR	12000
Qy	1201	PFLRNLSVSLVDLRGMLEERPARVRALBELLLGLIAGVFTPPPIATLPIARVADAFRSM	12600

Db	1201	FFLRNLSTSLVDLGNMLERARPARALLLELLGLIACGVFFPPPIATLPIRIVADAFRSM	1266		
Qy	1261	AAQAGLGLVLTLDGPEVOIRIPFHAGAPSTGDRDLDBLASAAPARAALAEFLRTQ	1320		
Db	1261	AAQAGLGLVLTLDGPEVOIRIPFHAGAPSTGDRDLDBLASAAPARAALAEFLRTQ	1320		
Qy	1321	VSQVLRTEIRKVGAAFLTRIGMSLMAVELRNTEASIKKUSTTFLSTSPNALLAQN	1380		
Db	1321	VSQVLRTEIRKVGAAFLTRIGMSLMAVELRNTEASIKKUSTTFLSTSPNALLAQN	1380		
Qy	1381	LLDALATLSLERRVAENLRAGVONDFVSSGADQDWEIHAL 1421			
Db	1381	LLDALATLSLERRVAENLRAGVONDFVSSGADQDWEIHAL 1421			
RESULT 2					
US-10-014-717-7					
Sequence 7, Application US/10014717					
Publication No. US20020192778A1					
GENERAL INFORMATION:					
APPLICANT: Schupp, Thomas					
APPLICANT: Ligon, James					
APPLICANT: Molnar, Istvan					
APPLICANT: Zitzke, Ross					
APPLICANT: Cyr, Devon					
APPLICANT: Goerlach, Joern					
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES					
FILE REFERENCE: 4-30582A					
CURRENT APPLICATION NUMBER: US/10/014,717					
PRIOR FILING DATE: 2001-11-13					
PRIOR APPLICATION NUMBER: US/09/335,409					
PRIOR FILING DATE: 1999-06-17					
NUMBER OF SEQ ID NOS: 30					
SOFTWARE: PatentIn Ver..2.0					
SEQ ID NO 7					
LENGTH: 2439					
TYPE: PRT					
ORGANISM: Sorangium cellulosum					
US-10-014-717-7					
Query Match	57.4%	Score 4137;	DB 14;	Length 2439;	
Best Local Similarity	43.4%;	Pred. No. 6.3e-309;			
Matches	933;	Conservative 135;	Mismatches 308;	Indels 776;	Gaps 13;
Qy	6	IERAAEDPIAIVGASCRLPQGVIDLSGFWTLLGSRPTVGVPAERMDAAAMPDPDAP	65		
Db	26	LEQRTETIAIVGICGRFGADTPPEAFWELLDGGRANV-QPDRRWALVGNPSEVP	83		
Qy	66	GKTPVTRASFISD-VACFDASFCGISPREALRMDPAHRLLEVCWEALENNAIAPSLVG	124		
Db	84	-----RWAGLLTEAVNDGFDAFFGTSPREARSLDPOORLLLEVWEGLEDAGIAPQSLDG	138		
Qy	125	TETGVFICIGSEXEALLPQATAEIDAHGCLTMSVAGRI SYALGRGPVAVDTA	184		
Db	139	SRTEFGIACGSDSYSHTVAAQORE-EDPADYITNTLSVAAGRI SYTLGLQGPCLVDTA	197		
Qy	185	YSSSLVAHLTAQSLRSGECSTALAGVSMISPTLVMLSKTPALARDGCKAFSAEAD	244		
Db	198	CSSSLVAHLTAQSLRSGECSTALAGVSMISPTLVMLSKTPALARDGCKAFSAEAD	257		
Qy	245	GFGRGEGCAVNVLRKLSGARADGDRILAVIRGSAINHDAASSGLTVPNGSSQETVLKRAL	304		
Db	258	GFVRGEGCAVNVLRKLSGARADGDRILAVIRGSAINHDAASSGLTVPNGSSQETVLKRAL	317		
Qy	305	ADAGCAASSVGVYEAHGTGTLGPPIEQALNAYYGLGRVAVATLLIGSVYTNLGHPEYA	364		
Db	318	QSAEVDAGALIGYVTHGTGTSLGPIEVEALRALVGPARDGASRCVGAAYTNLGHLEGA	377		
Qy	365	SGITGLKLVNLSLHGQIPAHILHAQALNPRTSMGDLRLVTTRATPPMDMTPRACVSS	424		
Db	378	AGVAGLICALALAHLELIRLHHTLNPRTIRIGTALALATEVVPMPRAGRPRFAEYSA	437		
Qy	425	FGMSGTAAHVLEAPARATCTPPAPEPPABELLVISARTASALDAQAARLDHLETTSPQC	484		


```

Db 438 FGLSGTIVHVVLEAPATVLA.PATPGRSABLLVLSASAAALDQAARLSNHTAAVEOG 497
Qy 485 LGDVAFLAATTSMEHRLAAVATSRBGLRALDAAAGOTS.PGAVSIAANSRGLAFL 544
Db 498 LGDVAFLVSTRSMERHLAAVATSRBGLSALVAAOGT.PAAGRAASGKLAFL 557
Qy 545 FTGOGAOTLWGRGLYDWSAFREAFDL.CYRLFNOELDRPLREYMAEPASVDALLDOT 604
Db 558 FAGGAGVPGMGRGLMEAMPAFRETFRDCTVTLFDRELHQPLCEYMMAPSGSRSLDQT 617
Qy 605 AFTOPALFTFYALAAALRWKGVPELVAGHSIGELVAAVCAVGFSLVEDAVFLVAAGRL 664
Db 618 AFTOPALFALEYAALFRSWGVPELVAGHSLGELVAAVCAVGFSLVEDAVFLVAAGRL 677
Qy 665 MQALPAGAMVSIAPEDVAAVAAPHAASVIAAVAPDOVYVAGAGOPHATIAAAMAA 724
Db 678 MQALPAGAMVSIAPEDVAAVAAPHAASVIAAVAPDOVYVAGAGOPHATIAAAMAA 737
Qy 725 RGARTKALHVSHPASPLMAEMLEAFGRVAVESVRSPIVLVSNLSGKACTDEVSAPGY 784
Db 738 RGARTKPLHVSHPASPLMDPMLAEPFRVTSVYRRPSIALVENLSGKPCCTDEVSAPGY 797
Qy 785 WVRHARAVVPADGVKALHAAAGACTFVEVGKSTLLGLVPA.CMPDARPALIASRAGDE 844
Db 798 WVRHARAVVPADGVKALHAAAGACTFVEVGKPTLLGLVPA.CLPDARPVLLPASRAGDE 857
Qy 845 PATYLEALGLMAVAGVSMAGLFPSCGRVPLPTYPMQRRYMI----DTKADAR-- 898
Db 858 AASALEALGCFVWGVSGVTSVSGFRVPLPTYPMQRRYMI.EAPVDREAGTGAR 917
Qy 899 -----PST----- 1292
Db 918 AGGHPLGVEVSSTHAGRLMETTLDKRL.PWIGERHAGEVVFPGAGYLEMALSAGAE 977
Qy 899 -----GDRAP----- 904
Db 978 ILDGPLOVDVLIETLTFAGDTAVPVQVVTTEERPGRL.FQVASEPGRBAPRIFA 1037
Qy 905 ----- 904
Db 1038 RGVLRIRIGVETPARSNLALRLARLHAAVPAALYVGALEWGLQY.GPALRGLAELMRGEG 1097
Qy 905 ----- 904
Db 1098 EALGRVNLPERAASATAYQLHPVLLDACVQWIVGAFADRDATMAFVGVSVALLFQSRP 1157
Qy 905 ----- 904
Db 1158 GELWCHARVVSDDGOASSRWGADPELMDGTGAVVAEISRLVVERLASGVRRRADDFLE 1217
Qy 905 -----GAGH----- 908
Db 1218 LDWEPAALGPKITAGRWLLIGEGGGLGRSLCSALKAGHVVHAAAGDDTSTAGMRALLA 1277
Qy 909 ----- 908
Db 1278 NAFDGAQPTAVVHLSSLDGGGQLGPGIGAQALDAPRSVDADALBESALMRGCDVSL 1337
Qy 909 ----- 908
Db 1338 VQALVGMDLRNPRLMLTRGAQAAAGDVSVQAPLLGLGRTIAL.EHAEILRCISVDLP 1397
Qy 909 -----DEVEGCAVRGGDRSARLDHPPSSGRRKREKAAAGDRPRLEI 952
Db 1398 AEPEGEADALLAEILLADAEVVALRGDRVLRVHL.PDAQRREKPEPGDPRFRIE 1457
Qy 953 DEPEVLHLVLRTERRAPGLGEVEIAVDAAGLSFNDVQLALGVNPPDDL.PKPNPPLILG 1012
Db 1458 DEPEVLHLVLRTERRAPGLGEVEIAVDAAGLSFNDVQLALGVNPPDDL.PKPNPPLILG 1012
Qy 1013 GECAGRIYAVGEGVNGVLVGOPIVIALSAGAFATHTTSSAALVLP.PPQALSAIEAAMPVA 1072

```

```

Db 1518 SECAGRIYAVGEGVNGVLVGOPIVIALSAGAFATHTTSSAALVLP.PPQALSAIEAAMPVA 1577
Qy 1073 YLFAMVYLDRIARLOPGERVLIHAATGVGLAAVQMAOHVGAETHAGTPEKRAYLESL 1132
Db 1578 YLFAMVYLDRIARLOPGERVLIHAATGVGLAAVQMAOHVGAETHAGTPEKRAYLESL 1637
Qy 1133 GVRVYSDRSRDFRVADVRAVMTGEGVDVVLNSLSGELIDKS.FNLLRSHGFVEIGKDCY 1192
Db 1638 GVRVYSDRSRDFRVADVRAVMTGEGVDVVLNSLSGELIDKS.FNLLRSHGFVEIGKDCY 1697
Qy 1193 ADNOLGRLPFLRNLSPSLVDLRGMMLERPARVALLLEELGLIAAGV----- 1239
Db 1698 ADNOLGRLPFLRNLSPSLVDLRGMMLERPARVALLLEELGLIAAGV----- 1289
Qy 1240 FTPEPIATLPIARVADFRSMAOHLGKVLTLTGDEPEVOIRI.PTHA----- 1286
Db 1758 LTPPEVTFPIISRAAEFRMAOHLGKVLTLTLDEPEVOIRI.PTHA----- 1817
Qy 1287 -----GAG----- 1289
Db 1818 GGLGGLGRVAGWLAERAGAGQLVLVGRSGAASAEQRAAVAL.EAHGARVTVAKADVDRS 1877
Qy 1290 ----- 1289
Db 1878 QIERVLRVETVAGMPLRGVVAAGLVDDGLMOQTPARFRTVNGPKVQALHHTLREA 1937
Qy 1290 ----- 1289
Db 1938 PLSEFVLVYASAGLFGSPGQGYAANAFLDALSHHRAAGLPALSIDKGMFTVGNVVA 1997
Qy 1290 -----PST----- 1292
Db 1998 QENRGARQISRMKGTTPDEGLSALARLLEGDRVQGVIPITPRQWVEFPATASRRLS 2057
Qy 1293 -----GDRDLIDRLASAPARAALAEFLTQVSQVLRTEIKVGAELF 1338
Db 2058 RLVTTORAVADRATGABRDLLEOLASAPASAGILQOVVQVSHVLRLEBDKIEVDAPL 2117
Qy 1339 TRLGMDSLMAVELRNRIEASLKLSTFLSTSPNIALLAONLL-DALATV 1389
Db 2118 SSKGMDSLMAVELRNRIEASLKLSTFLSTSPNIALLAONLL-DALATV 2169

RESULT 3.
US-10-014-717-5
: Sequence 5, Application US/10014717
: Publication No. US20020192778A1
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zirkle, Ross
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIIONES
: FILE REFERENCE: 4-30582A
: CURRENT FILING DATE: 2001-11-13
: PRIOR APPLICATION NUMBER: US/10/014,717
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 7257
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
US-10-014-717-5

Query Match 45.9%; Score 3311.5; DB 14; Length 7257;
Best Local Similarity 41.1%; Pred. No. 1,2e-24; Indels 531; Gaps 25;
Matches 784; Conservative 162; Mismatches 432;
Qy 9 AEDPIAIVGASGRLPGVIDLSGFWTLLEGSRDVTGVRPAERWDAAMPDPDAPGKT 68

```

```

Db 3021 ASDEIATVGAACRPGCVEDLESYMOQLAEGVVSABVPADRMADMDYDDPEIIPRT 3080
Qy 69 PYTRASFLSDVACPDASFFGISPREALRMDPAHLLLEVCWEALENMAIAPSAVGTETG 128
Db 3081 YTKAFAPRLDLORLDATEFRISPREAMSIDPQORLLEVSWEALESAGIAPDILLDSPTG 3140
Qy 129 VFIGGPEBEYEAALQOATASAEIDAHHGIGTMSVGAGISVALGRCVCVVDPAVSS 188
Db 3141 VFGVGPMEYTYQRLRGFTDGAAGLYGGGNMLSTYAGLSFTLGGFTLLMDTRCSS 3200
Qy 189 LVAVHIAOQSLSRGCESTALAGVSIMLSPTLWMLSKTRALARDGRCAFSAEADGFR 248
Db 3201 LVALHLAOCQSLRGECODALVGVVAVLAPETFTVLSRRLALSPGRCCTFEADADGYAR 3260
Qy 249 GEGCAVVVLKRLSGRADGDRILAVIRGSAIINHDASSGLTYPNGSSOEIVLKLADAG 308
Db 3261 GEGCAVVVLKRLRDORAGDSILALIRGSANVHDGSSGLTVPNGPAOQALRLQALSGAG 3320
Qy 309 CAASVGVYEAHGCTGTGDIPIEIOALNAVVGGRDVAATPLLIGSVKTMIGHREVASGIT 368
Db 3321 VSPVDVFECHGTGTALGDIPEVQALSEVYGPBSGDRPLVLAAGAKANAVHLEASGLA 3380
Qy 369 GLLKVLSTLOHGOIPAHHLAOCALNPRISKGLRLTVTRARTPMDMTPRRAGVSSFGMS 428
Db 3381 SLTKAVLALRHQIIPAOPELGELNHLPMNTLPVAVPRKAVPMGARGARRAGVSAFGLS 3440
Qy 429 GTNAHVLEEARATCTPPAPERPAELVLSAKTASALDAQAARLDHLETPSCQGDV 488
Db 3441 GTNVHVLEEARPEVAPAPAPRPELVLSAKSAAALDAQAARLSAHLASAPBELSLGV 3500
Qy 489 AFSLATTSAMHRILAVATSRGRLAALDAQAOCOTSPGAVRSIADSSRGTLAELFGQ 548
Db 3501 AFSLATTSAMHRILATTSREARLGAALDAQAOCOTPOGAVRGAVSSRGTLAELFGQ 3560
Qy 549 GAOTLGMGRGLYDVMSAFREAFDLCVRLFNCLEDRPLREVMVAEPASVDAALLDOTAFTQ 608
Db 3561 GAOMGMGRGLYETMPAFREAFDRCAVLFDRREIDQPLREVMVAAGLMAOALDGTAAQ 3620
Qy 609 PALFTEYALALMRSKGVPELVAGHSIGELVAAQVAFSLEDAVPLVAAKRLMOAL 668
Db 3621 PALFTEYALALMRSKGVPELVAGHSIGELVAAQVAFSLEDAVPLVAAKRLMOAL 3680
Qy 669 PAGGMVSIABEAPAVAAVAPHAASVSIATAVNAADDOVITAGAGPVHAAIAMAARGR 728
Db 3681 PAGGMVSIABEAPAVAAVAPHAATVSIATAVNGDAVITAGAEVOVTLGHTPARGR 3740
Qy 729 TKALVSHAFFHSPLMAFMLAFAGRVAESVYRPSIVLVSNLSGRACDDEVSSPGVYRH 788
Db 3741 TKRLAVSHAFFHSPLMDPMLEDFQVAAATIAVRAAPRVSVNTGVHAGPEIATPEYVYRH 3800
Qy 789 AREVVRFPDGVKALAAAGAGTVEVGPSTLLGLVPACMPDARPLLASSRGRDEPATV 848
Db 3801 VRSARVFPDGAALAAAGATVEVGPSTLLGLVPACIGEADAVLVPSLADRSECEV 3860
Qy 849 LEALGLMAVGLVSMAGLFPSSGRRVPLPTYPMQRERYIDTAKDA----- 896
Db 3861 LAALGAWYAWGALDMKGVFPDGAARVALLPMTPMQRERHMDLTPRSAPAPAGIAGRPLA 3920
Qy 897 ----- 896
Db 3921 GVGLMGPVAVLHVLSIGPRHQPLFGDHLVFGKVVPQAFHVAVILSTIAERMPRAEL 3980
Qy 897 -----ARGD-----RR----- 902
Db 3981 TGVEELKAIAMEPDEVEHIAVLTPEAAGDGLFEIATLAABETERMTTHARGVOPD 4040
Qy 903 -ARGA----- 906
Db 4041 GAFGALPRLLEVLEDRALQPLDRLSAVRIGKGLMRLWLDGRVGDASLTLVPT 4100
Qy 907 ---GHD-----EVEGGA-----YR-GGD 921

```

```

Db 4101 YPNADVAFLPILLDNGFAVSLSTRSEPEDDGPPLPFAVERVYMRBARVGRCCGV 4160
Qy 922 RRS-----RLDHP----- 932
Db 4161 PRSQAQVSSFLVDEGEVVAEVEGFCVCRAPREVLPROESGASTALYRLDWEAPLP 4220
Qy 933 -----ESGR 936
Db 4221 DAPARIESVWVAABGSEVAAALATRLNRCVLAEPKGLEAALAGVSPAGVICMEGA 4280
Qy 937 REKVEAAGR----- 946
Db 4281 HEEAPAAQVATGELSVQALRDRAVRLMWVWGAVAVEAGERQVATAVWGIGRTVM 4340
Qy 947 ----- 946
Db 4341 QERPELCTLVLEBEADARSADVLRLELGRADDETOVAFRSGRRVRLVKATTPEGL 4400
Qy 947 -----PERLIDEPVLDHLVLRTERBARGLVEIAYDAAGLSFNDVOLAGMVPDD 1000
Db 4401 LVPDABESTRLBAGOKTIDQRLAPQKRAQGPBEVEIKVTSGLNFTTILAVLGM----- 4456
Qy 1001 LPRKNPPLLIGSCACRIVAVGEGVNGLVVQGPVIALSAGAFTHVTSAAVLPRQA 1060
Db 4457 YPGDAGP---MGDPCAGVATVGGVHRVAVGDAMVTL---CTLHRFVTVDARLVVRQDAG 4511
Qy 1061 LSAIEAAMPVAYILTAWALRIARLOPGEVILHAATGVGLAAVQAOHVGAHVATA 1120
Db 4512 LTPQAAATVPAPAFITAMIALHDNLNLRGEVILHAAAGVGMAAVQIARVIGAEVPTA 4571
Qy 1121 GTPERKAYLESIGV---RVVSDRSRDRFVADVAVAMTGGEGVDVUNLSGELIDKSFNLR 1178
Db 4572 -SPSKMAAVQMGVPRTHIASRTLEFPAETROYTGGRGVUVUNALAGEVVDASISLS 4630
Qy 1179 SHGRFVELGKDCYADNO---LGLRPFRLNLSFSLVDLRGMLEBPARYVALLESLGLIA 1236
Db 4631 TGRFLFMKTD-IRDRAVAAAHGVRVFDILET-----APDRREILERVEGFA 4683
Qy 1237 AGVTFPPRIATLPLARVADARMSAQAHLKLVLTGLDPEVQIRIP-----HAGAP- 1290
Db 4684 AGHRALEPVAHFAITKEBAARFMAQAHOQKVL-LEPBAAPLAPGTVLLTGGLAL 4742
Qy 1291 -----STGDRDLDRLASAPARAAALBAPFRTQVSOVLTPEIKV 1333
Db 4743 GLHVARMLAOCQVHMVLTGRRG-LDTPGAAKVAIEIAGARVTIASVDADNNALEAV 4801
Qy 1334 ABAL-----FTRLGMSLMAVELNRRIEASLKLKLT 1365
Db 4802 LQATPAEMWPLQGVIIAAGALDDGVLDGQTTDRFSRVILAPKVTGAMNHELTAAGNDLAFV 4861
Qy 1366 TELSTS-----PNTALLAQNLIDLATLSLERVAAEVLRAQVQND 1406
Db 4862 LFSSMSGLSAGOSNTA-ANNTFLDUALAHRAEGLAOSLAWGPMWD 4909

```

RESULT 4
US-10-156-761-10436
; Sequence 10436, Application US/10156761
; Publication No. US20030119018A1

GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10436
LENGTH: 6146
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-10436

Query Match 32.1%; Score 2311.5; DB 15; Length 6146;
Best Local Similarity 32.3%; Pred. No. 9.5e-168; Indels 631; Gaps 37;
Matches 633; Conservative 178; Mismatches 519;

5 PIERAADPIAIVAGACRLPGVILDSGFMTLEGSRDYGRVPAER-WDAAMFDDPD 63
3983 PTAEAADDPVIVAMACRPFGEATSPKLMNDLIAAGDGGIDFPADRGMSVA-----D 4036

64 APGKTPVTRASFSDVACFDASFFGISPREALRMDPAHRLLEVCWALENAALAPBALV 123
4037 AARS---QTGGFLPDVAEPFAGFPGISPREALAMDPOORLLLETSMWALERAGVDALKLR 4093

124 GTENGVEFIGPSEYEALPQATSAEIDAHGCGTSPVAGAGISVALIGRGCVAVD 183
4094 GSRITGVVAGSHDYGTLTSLSGGODYALTGAVG---SVLSRIAYVLGLEGPALTVD 4150

184 AYSSSLVAHLACSLRSRGCSTALAGVSLMSPSTLVMLSKTRALARDGRCAFAEA 243
4151 ACSSSLVAHLAALAGLGEEDVALAGVAMATPDAFARCGGLAPGRCAFAADGA 4210

244 DGRGRGECVAVVILKRLSGARADGRILAVIRGSAIHHGASGLTVPNSSQEI VLKRA 303
4211 DGGWGGVGVVLVTLRSEARRRGRHDVLAVERGSANVSDGANSGLTLPNGPSQORVIRQA 4270

304 LADAGCAASSVGYEAHGTGTLGDPLEIOALNAVYGRVAPRLIGSKYKTMIGHPEY 363
4271 LASHGLSADVDVHAHGTGSLGDPLEAOALMTYQDRPADRPLWGSYKSYIGHTQA 4330

364 ASGITGLKTVLSLHOGQIPAHLLAQAOLNPRISGMDRLTVTRARTWPMDMTPRRAGVS 423
4331 AAGVAGVYKSVLARNGLPLRPLTHVDEPSREVDSAGAVELLTGRMPEPDRRAGVS 4390

424 SFGSGTNAVHVEEAPAA--TCTPAPERPABELLVISARTASLDDQAARLDHLETP 481
4391 AFGISGTNAVHVEEAPAA--TCTPAPERPABELLVISARTASLDDQAARLDHLETP 4449

482 SQCGDVAFSLATRSMEHRLVAATSREGLRAALLAAAGOCSPGAVSIADSSGKL 541
4450 DLTVADVGLSLAGTRAGLEHRAVLTGREDEFLVQLAALAEAGASAAAGVAGVA--GEGT 4507

542 AFLFTGGAGTGLGWRGLYDWSAFREAFD-LCYRLFNOELDRPLREVMMAEPASVPAAL 600
4508 AFLFTGGAGTGLGWRGLYDWSAFREAFD-LCYRLFNOELDRPLREVMMAEPASVPAAL 600

601 LDQTAFTOPALFTFEVALAALMRSGVPELVASHGISGELVAAACVAGVSLIEDAVFLVAA 660
4562 LDRVTFVQAGIFALEVALFELVSSWGRADVLTGHSIGELAAAYVAGVSLIADACRIVAA 4621

661 RGRMLQALPAGMAMVSEARPEADVAAVAAPHAASVSIAYVAPPOVITIAAGGVHAAIA 720
4622 RGRMLQALPAGMAMVSEARPEADVAAVAAPHAASVSIAYVAPPOVITIAAGGVHAAIA 720

721 AMARCARATLALVSHAFSPILMAPMLEAFGRVAVESVYRPSIVLVNSLGGKACTEVS 780
4678 TPAEQGRRTQOLAVSHAFSARMEMPLAEFAETILAAVEFRSPRPVPSVNTVGTGAAGAEFT 4737

781 SPGVTRVHAABEVAREADGVKALHAAGAGTVEVGPCKSTL-----LGLVPAQMPDA 830
4738 TPVYVWVHVEAARVAFADGVATVILARGVDRFELGPRGALTMAEETLDHGAADAVCPVL 4797

831 RPAULLASSRGRDEPATVLEALGGLMVGILVSAAGLFPSSGGRVPLPTPMQOSERWID 890
4798 HPE-----RPEDATLHLALATASVTGAADVMTTPP-ACARTTELPYAFQRRRTW-P 4847

891 TKADDAAGDRRAPG---AGH----- 908
4848 RPAASAGADLAAAGLAEGHPLLTAMLSPEGDVLTGRLSLATHPWLDGHTVLTGTVL 4907

909 ----- 908
4908 VPGTAFLVDLVGVGDRVCGGIRELTPLAPLAEBAVRLRVVLGAPDDTGCRPAVYA 4967

909 ----- 908
4968 QPDTDEGTWTHAAGLAPADSDTPAPRPBALDAMPVPGAEPVPMGFBGLADAG 5027

909 -----DEV-----EEGA----- 916
5028 FAYGPLFRGLRAARRGDEVEFAEVSPLPADSGAGFVGNPALDLAALHAGPAAQASDEPG 5087

917 -----VRGD----- 921
5088 SARLPSWGEVRVHAGADLLRVLRVLAADGTVTLDAADAGRPVVISGLVLRPLSPDR 5147

922 ----- 921
5148 LRAGTQAPDAPLAPATKIPLDVTGDTGDTARDTVECVLGEPSAGAMRSHPDAAVAAAL 5207

922 -----RSAR----- 926
5208 ADGKETPGVILACPLTGGPAAEARSAEWTLDLQWTLADRLTDSHLVIGTRHAAPV 5267

927 -----LDHPPESGREKEYBA----- 943
5268 ATPYAAPVDATDADAAFPVDAHGHEAPVDTAAALAOSTLAGLVSAQOTENPARITLVDPDT 5327

944 -----GDRPFLRIDEPGLD 959
5328 TAEDPAHLTAAYRGLGEVEVAVROGTLYARLTPREGRALTVPPEGEPMLDSTGRSLD 5387

960 HLVLVYTERRAPGL--GEVEIADVAGLSFNDVQLAGVVPDDLPGKRNPLLLGEGCAG 1017
5388 HLVLVPCPDAAALAGGTVAIVHAAGVNPRVDILALDM---YGRAD---LGTECAG 5439

1018 RIYAVGEGVNGLVVGGPVILASGAPATHVTTSAALVLPFPOLSAIEAAMAPAYITAM 1077
5440 VLEETGGVGLAGRWKMGVMAAGFPTAVADRVILARIIDMSFETAATVPVAFITAY 5499

1078 YALDRLARLOPGERVLIHAAITGVGLAAVOMAOHVGAEVATGTPREKAYLES-LGVRY 1136
5500 YGLVLDLGLGAGSVLVHAAAGGVMAAQLAHNLHAEVGTASBPKMSTLDTGLDRAH 5559

1137 VDSRSDRPVADYRAMTGGEGVDVNLISGELIDKSFNLLRSHGRFVELEGKDCVADNQ 1196
5560 IGSSRTLPADLVRETTGGAGVDVNLISGELIDKSFNLLRSHGRFVELEGKDCVADNQ 1196

1197 LGL-RPFLRNLSGLVDLRGMLEPARVALLLEELIGLIAAGVFTPPPLATLPIARVAD 1255
5620 VAADHGEVRYRSPDLGE-----AGPRLIAELIHLVLEPESGALTPLPVTTWDIRDATD 5673

1256 AFRSMAGQHLGLVTL----- 1273
5674 AFRVLSQATLTGKAALTVPAAGSVPEEAGETVLITGGTGLTGLLRLHVTGHEGRHYTLA 5733

1274 -----GDPEV-----QIRIPTHAGAPSTGDRDLDD-----RLASAPAPARA 1310
5734 GRGOTDPEVRQARADLAEMGAVERVA-CDADDEDAVRRLLBPLTANHRLAGVHAAV 5792

1311 ---AALEAPLRTVOVYLRTPELTKVGAELFTLGDNDLSMAVELRNRIEASLKXKSTTF 1367
5793 TDDGVVSLDRRLRSVLAH-PKYR-GAWNL-----HRLTHALPRMFLV 5835

1368 LSTSPNIALAON-----LIDALATALSLEVAENLRAGV 1403
5836 SSASATLTGAAGGVYAAAALFDLALAEHRAARGLPSSILAWGL 5878

RESULT 5
US-10-156-761-7962
Sequence 7962, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBATA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 7962
LENGTH: 6145
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-7962

Query Match 31.2%; Score 2249; DB 15; Length 6145;
Best Local Similarity 37.2%; Pred. No. 6.2e-163;
Matches 611; Conservative 183; Mismatches 545; Indels 304; Gaps 50;
QY 1 VADPRIEAED-PAIYAGSCRLPGVYIDSGFTLLGSRSDTGRVPAER-MDAAMF 58
DB 1531 VVEBOGHAADDEPAIYAMSGHLPFGVDSPEALMDLVTSGDAISEPFRGMDVELY 1590
QY 59 DDDPAPEKTEVTRASFSDVACFDASFEIGSPREALMDPAHRLLEVCWEALENNAIA 118
DB 1591 DPDDRPEKTYABDQFLYGATDFDAGFGISPRALAMPQORLMLTSEAFERAGIA 1650
QY 119 PSALVGTGTGFTIGGSEFEYALPQATASAEIDHGLGTMPSVGAIRISYALGRPC 178
DB 1651 PAORGRSTGVFVGMAVYOGYGVADVARTREG--VEGHRLVGGASVSGRVAYTFLEBPA 1708
QY 179 VAVDTAASSLVAVHACOSLRSGECSTALAGVSLMSPSTLWVLSKTRALARDGRCKA 238
DB 1709 VTITACSSSLVALHLMQSLRNGECTMALAGGVVMAGPSVFPEFSQKGLSPDKCKA 1768
QY 239 FSAEDGFRGEGCAVVVLKRLSGARADDRILAVIRGSAINHDGASGLTVPNGSSQEI 298
DB 1769 FGADADGTGMAEGVGMILVERLSDARRNGHQLAVVRSVAVNQGDSNGLTAPNGPAQR 1828
QY 239 VLKRALDAGCAASSVGVYEAHGTGTLGDPLEIOLANVYGLGHDVATPLLIGSVKTL 358
DB 1829 VIHOSLSAGLAPRDVDVAVEAHGTGTRLGDPLEIADALATYGGCGDGRPLMLGSLKSI 1888
QY 359 GHPEVASGITGLKLVLSLQGOIPAHLAQALNPRISM--GDRILVTRARTPMDMT 416
DB 1889 GHTQAAAGVAVIKVMAMRIGILPRTIHADDEPTPHVMSAGDVRL-ITEA-VDMPESDR 1946
QY 417 PRRAGVSSFGMSGTNAHVLEBPAPATCTPPAPERPAEL-LVLSARTASALDAQARLUD 475
DB 1947 PRRALISFGVSGTNAHTIIEQAPATAELPTPPDSCALVPMVLSGKGPALRAQAARLAD 2006
QY 476 HLETTPSQCLDVAFSLATTRSAMEHRLAVALAATSREGRALDAQAQOSTSPGAVRSIAD 535
DB 2007 HLDANDGTASDIGSLA-SRETFEDRAVLTLAOGSDKROALTALAAQOPANVTLGRA- 2064
QY 536 SSRGLAFLFTGGAGQTLGMRGLDVMSAFREAPD-LCVRLFENGLDRPLREVWMAEPA 594
DB 2065 RSEKGVGLFSGGQSORIGMRELYEAFPVFADAIDEVCAEL-----EAPV 2110
QY 595 SVDAALDQTAFTQALTFEYALALMRSWGVBEDEVAGHSIGELVACVAGVFSLEDA 654

DB 2111 DVDAETLHRTGCAQPALFAVEVALFRLLSGWVRPDYVAGHSVEIGIAAAHVAGVLSLDA 2170
QY 655 VETVAARGRLMOLPAGGAWVSEAPADVAAAAPHAA-SVSIAAVAPQOVVIAGAQ 713
DB 2171 VRLVSARPAALMOLPVGAMAVAVATEE---VLPHTBEVGIALINPSSMLSGAGE 2226
QY 714 PVHIAAAMARGARTALVHSHAFSPMLAPMLAEFGVAESVSYRPSIVLSNLGK 773
DB 2227 AVTMAIEVFGQGGKTRSLKVAHSHFPLMDPMLIEFAEVVRGLTFHQPVVSNLTGR 2286
QY 774 ACTDEVSSPGVWRHAEVVRPADGVKALHAAGTVEYVGPSTLLGLVPACMPDA--R 831
DB 2287 LA--EPYSPETWVRHVEAVRFADGIELTGELTTFVEIGPGGVLSGMVDDCVDEAVTV 2344
QY 832 PALLASRAGDEPATLEALGIMAVGVLSVAGLPPSGGRVPLPTYPQORRYMTD 891
DB 2345 PVL---RGDRPEQAAVVTALAEHLHGVSPDWALPP-GAHRVTLPTVAFQREPFMDG 2399
QY 892 KADDAAG--DRAPGAGHDEVE---EGGAVRGDRRSARLDH--PPESGREKYE-- 941
DB 2400 EGDLAAGTPTDADSGFMDSVREDAESLAATLGVSADASLGAILPRLSMWRQRRQS 2459
QY 942 -----AAGDRPF---RLETDEPGVLDHLVLYTER 968
DB 2460 VVDMRWYVTKPPLGALPQPSAGSTWLLVVAESEMVAASYRTALDERGL--ELVTLVA-- 2515
QY 969 RAPGLGEVETAVDAGAGLSPNDVOLALGMVPPDDLCKENPPLLGSECGAGRIVAVGEGVNG 1028
DB 2516 -GPTDRTGLRLAEAGV--GPVGTVLSLDEDSASAGDQLSGLAN--TLCTVQALGD 2570
QY 1029 LVWGQPVIALSAGAFATHVT-----TSAALVPR-----POLSAIE 1065
DB 2571 VGVDPPLMCAIRGAVAIAGRSRVDRLPLOSQWVGFRAPAALEHPRMGGLDLPELDART 2630
QY 1066 AAAMPVAVLTAWALDRIA-----RL-----OPGERVLIHAATGVYGL 1103
DB 2631 AAR--VAAVIGQTAEQDVAVASGVLRRLVRAARATGPAQGSPRGTVLITGTGALGG 2688
QY 1104 AAVQWAGHVGAEVH-----ATAGTPEKRAYLESIGVR----- 1135
DB 2689 HVARMLAGAGL-HLVITSRRGPAAPGAELVLAELGAAATVAVCDADRDLRALLLA 2747
QY 1136 -----YVSDRSRDFVADVA----- 1151
DB 2748 QHPVNAVHAAVGDHMIEDSDPAGFAGVAAAGAAATHLDELLAGOELDAFVMPSSGA 2807
QY 1152 --WTG--GEGVDVVLNSLSGELIDKSFNLRSRGR-----FVELGRDCTVADNQLGL 1199
DB 2808 GIMWGAGQAGVSAANAVILDLAEHR---RAHGRTALAVSWGVAEGMAG-VGDGCEML 2862
QY 1200 RPLRLNLSFSLVDRGMMLBRPAVRALBELL--GLIAGV-----FTPPPIATLPI 1250
DB 2863 R-----RRGLPRMKPPLAISALQOALDGETALTVDPMERFICGFTVTRPS 2910
QY 1251 ARVDAFRMAQAOHLKLVLTLDPEVOIRIPHAAGSPSTGRDULDLRLASAAAPARA 1310
DB 2911 ALFSDV-----PEAGRALAEPAAGAAAGNA-LAAKLTGLPAABQD 2950
QY 1311 AALEAFRLTOVSQVL---RTPETVAGABALFTRLGMSLMAVEIRLIESLKLSTTF 1367
DB 2951 RTLVDLVTRTHAAVLTGHDGTAAPVPG--RAFENDIGFDSLAVELRNKLTADTGKLP TTL 3008
QY 1368 LSTSPNIALLAQ---NLDLALATLALSERVAE-----NLBAGVON-----D 1406
DB 3009 VEDYPNATLARFLRAELLSRGAARQKTAAGDEPIALVMSCLPGGVSGSPEDLMQ 3068
QY 1407 FVSSGA-----DQDMETIAL 1421
DB 3069 LVTSIGDVISGFPEPDGRMVEAL 3091

RESULT 6
US-10-014-717-6

```

; Sequence 6, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Ziklie, Rose
; APPLICANT: Cyr, Devon
; APPLICANT: Geerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3798
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-10-014-717-6

Query Match      31.1%; Score 2245.5; DB 14; Length 3798;
Best Local Similarity 36.2%; Pred. No. 5.6e-163;
Matches 605; Conservative 191; Mismatches 495; Indels 379; Gaps 43;

QY 9 AADDPAAIVGASGRPGGVIDLSGFWTLFSGSRDTGVRVPAERMDAAAFDPDPDAPGKT 68
DB 1519 AADDDAIVGACRPFEGDEGLTMYRHLAEGNVSTVEPADWRRAADMDVDPPEVGGT 1578
QY 69 PYTRASFSDVACFDSFPGISPREALRMDPAHRLLEVCMEALENNAIPASLVGTETG 128
DB 1579 YVAKGAFLRVRSLDAFAAISPREMSLDPQGRLLLEVMSEAIERRGDPMLRBSATG 1638
QY 129 VFIGIGBSEYEAALPOATASAEIDAHGGLCTMPBSVGAGRISVALGRCPCVAADTVAYSS 188
DB 1639 VFQGMIGSEHAERVOGLDDDAAL-LYGTGNLTVSAAGRSLPFGILGHPMTVDYACSSS 1697
QY 189 LVAVNHACOSLRSGECTALAGVSLMSPTLVMYKSTKALARDGCKAFSAEADGFR 248
DB 1698 LVNLHLACOSLRJGECQALAGSSVLSRPSFPAASRMRLLSPDCKTFSSAADGPAR 1757
QY 249 GEGCAVNVLRKLSGADRADRIIVIRGSAIINHDCASSGGLTVPNGSSQEIIVLRKALADAG 308
DB 1758 AEGCAVNVLRKLRDAQGRDPIILAVRSTAIINHDCGSLTVBSGRQCALRQALAAQ 1817
QY 309 CAASSGVYERAGGTGTTGDPPIEIQALNAVYGLGRDVATPLIGSVKTNLGHPEYASGIT 368
DB 1818 VADAEVDVEFCHGTGALGDPPIEQALGAVYGRGRPAERPLMIGAVKANLGHLEAAAGLA 1877
QY 369 GILKVVLSLQHQIIPALHLAQAALNPRISMGDDLTVTRARTPMDMTPRRAQVSSGMS 428
DB 1878 GVLKVLTLAHEHDIPOPELDELNPHI PMAEPLPAVVRRAVPWRPGARPRAGSARGLS 1937
QY 429 GTNAHVLEBAPATCTPPAPERPAELIVASATASALDAQAARLRDHLTETYSQCLGDV 488
DB 1938 GTNAHVLEBAPAVEPVAAPPERAAELFVLSAKSAAALDAQAARLRDHLTETYSQCLGDV 1997
QY 489 AFSLATTRSAMEHRLAVAAATSRGLRAALDAQAAGQTSPGAIVSIAD-SSRGKLAFLFTG 547
DB 1998 AFSLATTRSAMEHRLAAVAAASRELRGALSAAGQHTPGAVGRGASGAPKVFVFP 2057
QY 548 QGAGOTLGMGRGLVDVMSAFREAPDLCYRLFNQELDRDLREVMAAPRSVDAALLDQAF 607
DB 2058 QGQOWMGWKMLAEERVPFAALEGCDRAI EAEGMSLIGELSDXA---ASQGRDIV 2114
QY 608 QPALFTFEYVLAALWRSNGVEPELVASHISIGELVAACVAGVSLFEDVFLVAARGRLMA 667
DB 2115 QPVLFAVEVALSLMRSGVPEAVVGHSGVEVAALVAGALSIEDVVAIICRSRLRR 2174
QY 668 LPAAGANVSIADPADVAAA VAPPAASVSI AAVNAPQOVVIAGAGQPVHAI AAMAARGA 727

```

```

DB 2175 ISGGEMALVEISLEBAEALRGHEGRSLVAVANSNPRSTVLAGEPAALBEVLAALTRAKV 2234
QY 728 RTVALHVSASFHSPLMAPMLAEFGVAESVSYPSPISVLVSNLGRACKTDEVSSPGYWR 787
DB 2235 FMRQVAVDVASHSPQDPLREBELIALGAI RPPAAVPMRSTVTGVIAGPELGAYMD 2294
QY 788 HAREVVRPADGVKALHAAGTFFVEVGPSTLGLVPACMPDAP--ALLASRAGDEP 845
DB 2295 NLRQVRFAPAAAQALLLEGCPALFIEMS PHPIIVPPDEIQTAAEQGAAGVSIIRRCODER 2354
QY 846 ATYLEALGLMAVGVLSWAGLPPSGRRVPLPTYMORERYIDTKADAAAGDRAPG 905
DB 2355 ATLEALGLTLMASGVVSWARLFPAGGRVPLPTYPQHERYWI---EDSVHSGK--- 2406
QY 906 AGHDEVEEGAVAGGDRRARLDHP----- 930
DB 2407 -----PSLRRLRLRGATDHPILGAPPLVARSAPGALHMQALSDERLSYLSHRVH 2457
QY 931 -----PPESGRREKVEAAGDRPFRLEIDEFG-----VIDHLVL-----RVTE--- 967
DB 2458 GEAVLPSAAVVEWALAAQ-----VDLYGTATLVLEQALERALAVPSEGRIVQVALS 2510
QY 968 RRAPRGVEVIAV-DAAGLSFNDVOALGMV-----PDDLPGK 1004
DB 2511 EEPGGAASFOVSSREBAGRSW--VRHATGVCSGSSAVALKEAPWEIQRCPVLSSE 2358
QY 1005 PNPPLT-----LG-GECAGRIVAGGVNGLVVGQPVIALSAGAFAT 1045
DB 2569 ALYPLNHAALDYPCFGQVEQWLTGTGEVLRVRLPGD-----MASSGAYRI 2617
QY 1046 H-----VTTSAALV-----LPRP-----QALSAI----- 1064
DB 2618 HPALDLACQVLTALITTPESIERRLRDLHEPDLPRSRAPVNOAVSDTWLMDAALDGC 2677
QY 1065 --EAAAMPV-----ALTAMVALDRJAR-----LQGERVLIHAATGCVGLA 1104
DB 2678 RRGASAVPVLDLVGSHFAKMEVWERLAQAYIIGTLRIWNVFCAAGRHITDELIVRQIS 2737
QY 1105 AV-----OMAOH-----VGAEVHATAG--TPKRAYLESIGRVYSDSRDFVA 1147
DB 2738 VYTRKVIKRMELVAIGLVGDGEHFVSSQPLPEPDLAAVLEAG-----RVFA 2787
QY 1148 DVRA--WT--GEGVDVVL-----NSLSEGLI 1170
DB 2788 DLVPLFEMCKFAGERLADVLTKTLALBELPFGSSFDMAERIVRDSPIARYSNGIYRGV 2847
QY 1171 DKSFNLRSHGR--VELG-----KRDY-----ADNQLGLR 1200
DB 2848 ESAPRVVAVSGMFSILEIGATGATTAAYLVPLLPRTEYHFTDVSPLFLARAEQRRDY 2907
QY 1201 PPLRNLFSFLVD-----LRGMLEPRARVRLBELLGLTAAGVF-- 1240
DB 2908 PFLK---YGLVDVQEPAGOGTAHQRFVIVAAVNHATRDIAATKRLSLSLAPGLLV 2964
QY 1241 ----TPPI-----ATLPIARVADAFRSMAOQH-----GK 1268
DB 2965 LVEGTGHPIMFIDITGLIGEMQYEDDLRIDHLLPARTWCVDVLRVGFADAVSLPDGGS 3024
QY 1269 LVITLGDPEVQIIRPHAGAG-PTGDRDLDRLLAAPAPAAA-----A 1312
DB 3025 PACILIOHVILSAPPIAGACDSGES-----ATESPAAARVROEMWADGADVHRMA 3078
QY 1313 LEA--FLRTQVSOVLRTPEIKVGAELPTRLIGDSIMAAVELRRIEASLK 1360
DB 3079 LERMTYHRPRGRQVWVHGRLRTGGG-FTKALAGDLLLEFDTCQVAAEVO 3127

```

```

RESULT 7
US-10-156-761-7961
; Sequence 7961, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI

```

APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 7961
 LENGTH: 3352
 TYPE: PR
 ORGANISM: Streptomyces avermitilis
 US-10-156-761-7961

Query Match 30.5%; Score 2199.5; DB 15; Length 3352;
 Best Local Similarity 37.2%; Pred. No. 1.6e-159;
 Matches 611; Conservative 182; Mismatches 529; Indels 321; Gaps 51;

4 PTERAEDPRAIVGASGRLGCVLDLSEFTLLEGSFDTYGRVPAER-NDAAMPDDP 62
 27 REVEKDDPFAIVGMACRYGVRSPEDLMLVAAGDAVDPADGMVEGIDYDDP 86
 63 DAPKTPVTRASFSDVACFDASFFGISPREALNMDPAHRLLEVCMEALENAAIAPSL 122
 87 DAKRTYARRGGFLYEAGEFPDPPFGISPREAVAMDPOGRLLLETTWTTFERAGIDAESV 146
 123 VGTETGVFIGGSEYEALPQATASAEIDAHGGLTGPVSGAGRIYALGIRPCVAVD 182
 147 RGSRTGVFGSGYDYMVL--VGVGESDHLGTGNASVMSGRIATFGLGPAVTVD 204
 183 TAYSSIVAVHLAQSLSGSCSTLACVSLMSPSLVWLSKTRALARGCRAFAE 242
 205 TACSSSLALHMAIQALNNGECSMALAGVOVMTPTFAVERSRGRGLAPGRCAFGAG 264
 243 ADGFGREGCAVVVLKRLSGARADGRIILAVIRGSAINHDGASSGLTVPNGSGOEIVLKR 302
 265 ADGGMAGVGMLVERISDAVRNGHEVLAVRGSANVDGASNGLTAPNGPAQORVIRQ 324
 303 ALADGCAASSGVYEAHGTGTTGADPIEIQALNAVGLGSDVATPLLIGSVKTNLHPE 362
 325 ALASAGLSPAQIDVVEAHGTGTTGADPIEQAALLATYGRERTEDRPLMLGSLKSNIGHSQ 384
 363 YASGITGLKVVLSLQHQIIPAHLAQALNPRIS--GDLRLTVTRATPPMDMNTPRRA 420
 385 AAGAGGIIKVMAMRHGVLPTLHVDEPTPHVDSAGAVRL-LTEA-VEPETHPRRA 442
 421 GVSSFGMSGTAHVLEBPAPATCTPPAPERPAELL-----VLSARTASALDAQARL 473
 443 AVSFGVSGTAHHTIEQAF-----PAQDEGAGLVDPGLAMVLSASDALARRQARL 497
 474 RDHL--ETYPQCLGADVAFSLATRSAMEHRLAVALTSREGRLAALDAAGQTSPEGAVR 531
 498 RPSARGEVRPQ-----DIGLSLATTRAAMRHRAAVVGDREELSLGLEDLATG--TPPSRV 551
 532 SIASSRSKCLAFLEFGAQTLLGMGRGLYDVMSAREAFD-LCVRLTFNQEIDRLAREVM 590
 552 LLGRPAGKGTFLPSGGSORIGMRELYAAYPVFAAAYDEVCAHL----- 597
 591 AEPASVDALDQAFQPAFTFEYALALMRSGVPELVAGHSIGELVAACVAGVFS 650
 598 DAPVDVDETHLRICGTOPALFAVEVALFRLLSEIGVRPPDVGHSVEIAAAHVAGLS 657
 651 LEDAVFLVAARGRLMQALPAGCAWVSIAPADVAALVAAPHAA-SVSIAAVNADQVIA 709
 658 LDDAKIVSABAAIMQALPAGGAMVAVOATEE-----VLPHLTDVSIAAVNGPSSVVS 713

710 GAGPVHAIAAAMAARGATKALVSHAFSPMLAEMLEAFGRVAESVSRPSIVLSN 769
 714 GDETAALAIAGFBAEQKTSRLKVSHAFSPMLDPMLEAFVVRGLAEQOPLPVSN 773
 770 LSGRACTDEVSSPGYVWRHAREVRFPADGVKALHAPAGATFEVEGPKSTLLGLVPACMPD 829
 774 LTQGPV--EAVTPYVWRHAREVRFPADSVRTLHDLGVKTFIEIGPGVLSGMAQCADD 831
 830 --AAPALLASSRACRDPATVLEALGSLMAVGLVSNAGLPPSGGRVPLPTYWQREY 887
 832 IVTVPLV--RADRPERRQAVVTLAHLHTGVSVDHMSFF-AGARTDLPYAFQREMY 886
 888 WIDTKADDA-----RGDRRAPGAGHDEVEGAGVARGDRSARLDHPPE 933
 887 WIDAPESAAAEADPVDAEFWEYERGLQA-----LAETLIDGADDTGDLV--PRLS 937
 934 SGRREKVE--AAGDRPRL-----EIDEPGLDHLVRYTERRAPALGEVEIAVDAAGL 985
 938 SMRQRKEQSTVDDRVRRESWKRLGELGPAIGGTWLLAVPAEENQTAARVTRALARGA 997
 986 SPNDVOLAGMVPDDLFGKRNPPILLGECAGRIVANGEGVNGLVV-GQPI----- 1036
 998 TLKT-----LVDDAASHRAG--LAGELAG--IGPVDGLSLVNGDPLVPTLLVQA 1045
 1037 -----ALSAGAFATH-----VTSAAVLPR-----POL- 1061
 1046 LGDAGVDAPLMCLTSGVAVVSGSDAVDARHAQVWGIRVALLPRMKGLDLPESVD 1105
 1062 --SAIEAAMPVAVLTMAVLAIRIA-----BLQGERVLIHAAT 1098
 1106 DQAAVRLADVLLEQLTGRWEEDQLAVASGVFARLHAAPARAATRMQGRGTVLINGCT 1165
 1099 GGVGLAIVQNAQHVGAENH-----TAGTPEKAYLESIGVR--YVSDNSDR--- 1144
 1166 GALGHAARMLARGAE-HLVTSRRGADAPGAALRDELEVIGARVTFACDVADDAV 1224
 1145 -----FVADVAMTGGEGVDVVLNSGELIDKS 1173
 1225 AALLAQHFTFAVHAAGVADAGVDTATTPAFAAALAAKYGGA--AHIDELLDQDELDA 1281
 1174 FNLLSHGRFVELGRDCVADNOLGRLPRLNLSFSLVLRGMMLERPARVRLLELL- 1232
 1282 FVLFSISGVWSSGQAAYAG--NAFLDGLA-----RQRDRGILTATAVS 1325
 1233 -----GLIAG-----VETPPRIATLPARVADARSMQAQHLKVLTL 1273
 1326 WGPVADGMAVADGDDEERLRRGLRAMTP-----ASATLSLQRLADRDELTLTV 1374
 1274 GD-----PEVO--IRIPTHAGSPSTGDRDLRLASAPAPAR 1309
 1375 ADVVMSRFIVPFTLGRSPPLGDLPEVRALAVDTPA-AHGSG--ALAESIALSHQDR 1431
 1310 AALAEFLRQVSGVL-RTBEIKVGAELFTRLGMSLMAVELNRIEASLKLTSTFTL 1368
 1432 ABALVDVTRTHAAVVLGHRGAGAVEADRPREDIGFDSLTAVELNRIKATGTSLPATLI 1491
 1369 STSPNILLANONLLDALATALSIERVAENLRA-----GVON-----D 1406
 1492 FDHFTADLAAHLAEI--TGGQHEAVAAATAVLALEPIAIIIGMACRYPGGVRSPELME 1550
 1407 FVSSGAD-----ODMEIILAL 1421
 1551 LVASGRDAISRFPGRNGVDEAL 1573

RESULT 8
 US-10-156-761-7964
 Sequence 7964, Application US/10156761
 Publication No. US20030119018A1
 GENERAL INFORMATION:
 APPLICANT: OMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHIYUKI
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO: 7964
 LENGTH: 3564
 TYPE: PR1
 ORGANISM: Streptomyces avermitilis
 US-10-156-761-7964

Query Match 30.5%; Score 2197.5; DB 15; Length 3564;
 Best Local Similarity 34.7%; Pred. No. 2.5e-159;
 Matches 638; Conservative 187; Mismatches 546; Indels 467; Gaps 57;

QY 4 RPIRAAEDPIAIVGASCRPLPGVITDLSGFWTLLEGSDTVGRVPAER-WDAAMFDDP 62
 DB 26 RDVEDADAGEPAIIIGMSCRPGVDSPEDELMELVAGRDVAEFPADRGMDLDALYDDP 85
 QY 63 DAPKPTPTRAFSDVACPDASPRGTSPREALNMDPHRLLEFCWCEALNMAIAPASL 122
 DB 86 ESRKTSYARBGAFYDLRPAAGLFGISPREALANDPOORLLETSMLEFERRAGIDDAASL 145
 QY 123 VGTETGVFTIGIPSEYEAL--PQATASAEIDAHGIGTWPVSGAGRIAYALGIRGCV 179
 DB 146 AGSTGVAGVWNEHFLALQNSPQ---DLEGVLGTSGSVASGRVATPFGLEGAV 200
 QY 180 AVDTAYSSLYAVHILACQSLSGECSTYLAGVSLMSPTLWLSKTRALARDRCACF 239
 DB 201 TVDTACSSLYAVHILAVQSLRNGECSILVAGGVMTVPATFVGRGRLAPDGRCKP 260
 QY 240 SAAADGFERGRCVAVVVKLRSARADGDRILAVIRGSAINHDAAGSLTVPNGSSORIV 299
 DB 261 AAGADGTGMBGAGLLVERLSADARRNGHEVLAVRGSANVODGASNGLTAPNGSQQRV 320
 QY 300 LKRALADGCAASSVGVYEAHGTGTTGDPTEIQALNAVYGLGRDVATPLIGSVKTYLG 359
 DB 321 IRQLAAQAQLTTAGVDVAEAGHTGRGDPTEAQLLATYQGRPDOKPWLGSIKSVIG 380
 QY 360 HPEYASGTTGLKLVLSIQGQIPAHILHAQALNPRISV--GDLRLTVTRASTPMDWNT 417
 DB 381 HTQAAAGVAGIIKVMAMRHGELPRTLHVDAPSVHVDTAGAVRLLT--ERTPMSASDP 438
 QY 418 RRACVSSGMSGTAAHVLEAPAA-----TCTPAPERPALL----- 456
 DB 439 RRAGISSGSGTAAHTTIEEAPSAEAPQDTGAPEGEBEGESAEPVSAEAPPAAGP 498
 QY 457 ---VLSARTASALDAQARL--RDHLETPESQCLGDVAFSLATRSAMEHRLAVALTSRE 511
 DB 499 VPMWISGTEERLAQAERLLSRDPAEYPL----DVAFSLATRTALEHNAVLOGETPA 554
 QY 512 GLRALDAAAOQOTSPEAVRSIADSSRGLAFLEFGAQTLGMRGLYDVMSAFREAFD 571
 DB 555 ELAAGLRAVAGSPARVARGLSGTG--GRVGFLEFGQSQRLGMGRRELYASYPAAAYD 613
 QY 572 -LCYRLFNGELDRPLREVMAPASVDALLDQRAFPTOPALFTEVYALAAAMRSWGVPE 630
 DB 614 EVCAQL-----DAVVDVSEELHRTGAAPALFAVEVALFLLSESGRAD 659
 QY 631 LVAGSIGELVAACVAGVSLIEDAVFLVAAGRLMQALPAGAMVSIAPADVAAYAP 690
 DB 660 YVAGHSVGEELAAHVAGVSLDDAKLVASARALMQALPAGAMVAVAVQATDE----VLP 715
 QY 691 HAA--SVSIAAVNADQVITAGGQFPAHIAAAMAAAGARTALHVSASFPLAAMLEA 749

DB 716 HLTDAVGAIAINGRVSVAEADVATAIAEAFRQGRKSSRLKVSHAFSLMDPILIE 775
 QY 750 FGRVAESVYRPSIVLVSNLGCACTDEVSPPCYWVRHAREVRFADGVYALHAAGCT 809
 DB 776 FAEVYRSLTEHRPOIPVSNLTGTLA--EPTTPQYWRHREAVRFADGVYTLNDLGVT 833
 QY 810 FVEVPSKTLGLVPACMPD--ARPALLSSRAGDEPATVLEALGGLMAVGLSVWAG 866
 DB 834 FVEIGPGVLSALAEGLDDDVTVPAL-----RADREERVAVVAESELHAGVSPWHA 889
 QY 867 LFPSGGRVPLPTPMQREYRITKADDAAGRRAPAG--HDEVEBEGAGRGGR 923
 DB 890 FFP--GAARRVDLPYAFQERYWLD--ASPAPGGVRAAGLSADHPLGAASVLAGDER 946
 QY 924 --SARLD--HP-----PES 934
 DB 947 LHTRLSLRTHRLMSDIAVLTALLPGTAPVELAVRAADENGALLLEDLTLEALVVPQ 1006
 QY 935 GRRE-----KVEAAGRPRL-----EIDEPVLHLVLTERRA----- 970
 DB 1007 GGVAVQVWVAGVADGTGRPLTVHSRPEDSDLPVW--RHATGVTEGSEBAGAAADRLLT 1065
 QY 971 --PGLGEVETAVDAAGLSFNDVOLALGM-----VPD--DLP 1002
 DB 1066 AMPPTGSEPYDLD-----GFYDRLAALGLAYGPAFRGLRSAMRTGEDVFAEVALPDGDTG 1121
 QY 1003 GKPNPRL-----GGEAGRIYAVGEG----- 1025
 DB 1122 SFLHPALLDAAHATGAGAGDSLVABADGPLLFPAMSRVSATGASTLRVLRAPGT 1181
 QY 1026 --VNGLV--VGQVVI-----ALSA-----GAPAT 1045
 DB 1182 DAVSLVADAGEBVAVESITLPAVSAERLKRSGDAMFTVERAPITLPSDDADGTA 1241
 QY 1046 HTTSAAL-----VLPFQA-----LSAIEAAMPAYILT--AMYALDRITRL 1086
 DB 1242 YVPLAALAEADGSPQDVVVPCCPDEGATEAERAAVSTDVLRVQRFADRPRL 1301
 QY 1087 QPGERV--LHHAATGVGLAVQWAGVGAEVHATAGTPEKRAYL----- 1129
 DB 1302 VLMARCDLHAHAAAGLVRA--QAEINPGRVLTLETFRPDBAALVGCYVLSGPHVYR 1359
 QY 1130 -----ESLGVRYVDSRSRDFVADVRAMTGEQGVVYVLSLGEILDKSF--N 1175
 DB 1360 EGEVASRRLAARASLATEADVATGRADEMASALVGAASRLGTVLTGAAGL--GKTLARH 1418
 QY 1176 LMRSHG-----RVELEKRCICYADNOGLRPLRLNSF----- 1208
 DB 1419 LVYTGVRRLVLSRRGADAPGADLKAELAAAGAEATWACDLAERDALRLAATPVD 1478
 QY 1209 SLVDLRGMLE-----RPARVRAL-----LEELG-----LIAGVFTPP 1244
 DB 1479 SVHTAGVLDGVTALTPERVGAVLPAKADAVLNLDELGTSPSTFVLFSSAAGVFGNG 1538
 QY 1245 IATLPIARV--ADAFRSNAQO----- 1264
 DB 1539 QGNVAANAFLDAPARBRHQAQRTVSLANGLWEGQAMADABDEAGSRMAAGVLP 1598
 QY 1265 -----HUKVLVTIGDPE--VOIRIPTHA-----GAGSTGDRD 1236
 DB 1599 TEDGLRFDAALSNBEVLPVVRDLTAALNRRAETVPALRLGLVRSRLRTAORAAAGS 1658
 QY 1297 LUDLSAAPARAALAEFLRTOVSOL--RTPEIKVGAELFTRLGMSLMAVELNRI 1355
 DB 1659 LAERLAGIADDERBALLEVRGEVAAVLGHSSRAVOPTYAFODLGFDSLTAVLRRL 1718
 QY 1356 EASIKLKTSTFLSTSPNIALAONT-----DALATALSERVAAE----- 1397
 DB 1719 ATATGALPLPVLVDHTPALAAHIGTELLEGEAPATYABAAARQTAADPEIAYGMS 1778
 QY 1398 -NLRAGVQND-----FVSSGAD-----QWETIAL 1421
 DB 1779 CRLPQGVRSPEDELMRLVASGRDGTTRPPDRMGVDGL 1816

```

RESULT 9
US-10-156-761-7965
; Sequence 7965, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7965
; LENGTH: 7746
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7965

Query Match      30.0%; Score 2159.5; DB 15; Length 7746;
Best Local Similarity 35.0%; Pred. No. 6.8e-156;
Matches 626; Conservative 181; Mismatches 532; Indels 451; Gaps 56;

QY 11 EDPALVAGACRLPGGVIDLGSFWLLBGSRTDVGKPAER-WDAAMFDDPPDAPGKTP 69
DB 5971 DEAIIVMACRFPDGVSPEDLMRLVASGDAISGFPEDRGMDEVENIYDPPDRSGKSY 6030
QY 70 VTRASFLDVACFDASFCISPREALRMDPAHRLLEVCWEALNMAAIAASLVNTGTCV 129
DB 6031 VRHGFLLEAEFDDAFPGISPREALMDPQRLLETSSWEAFERAGIDPAIVRGSRTGV 6090
QY 130 FIGIPSEYEALPQATASAEIDAAGLGCTMPVAGARISYALGRCPCVAVDTYSSSL 189
DB 6091 FAGWYHYGRV-KTAPEGMDAYLGSSAGSISGRVSTFGLEGRAVYDTACSSSL 6148
QY 190 VAVHLACSLRSGECSTALAGVSLMLSPSTLVMLSKTRALARDRCFAFSADGFGRG 249
DB 6149 VALHLATOLRSGECSMALVGVTVMAFSTFEFSRQGLSADRCFAFAGADGTGWA 6208
QY 250 EGCAVAVLKRLSGABADDRIIAVIRGSAINHDSASGLTVPNGSSQELVLRALADGC 309
DB 6209 EGAGMLVERLSDARNGHTVIAVVRGTAVNODGASNGITLAPNGSQQQVIRQALANGV 6268
QY 310 AASGVYEAGTGTLLGDPLEIQALNNAVYGLGRVATPLLGISYKINLGHEVASGITYG 369
DB 6269 SSEQVDAVEAHGTGKLDPIEAQALLIAFYGERPGDRPLWGLSKNSINIGHAQAAGVGG 6328
QY 370 LKVVLSIQHGOIPALHQAOLNPRISW-GDLRLTVTRARTPPWDMNTPRRAGVSSFGM 427
DB 6329 VIKWMAWRHGALEPRTLIVDEPTPHVDWSAGDVR-LTEA-MEWETHTPRAAASVSGEI 6386
QY 428 SGTNHHVLEBAPATCTPPAPER-PALILVLSARTASALDAQARLMDHLETYSQCL 485
DB 6387 SGTNHHVLEBAPATCTPPAPER-PALILVLSARTASALDAQARLMDHLETYSQCL 6443
QY 486 -GDVAFSLATTSAMEHRLAVALATSRREGIRALDAAGQOTS-----PGAVERSIADSS 537
DB 6444 PADIFSLATTSAMEHRLAVALATSRREGIRALDAAGQOTS-----PGAVERSIADSS 6495
QY 538 RGLKAFLEFTGGAQTGNGRGLYDVWSAFREAFDLCVRLFNOELDRPLREVMMAEPASVD 597
DB 6496 GKSQGFLESGGSGRITGNGRGLYDVWSAFREAFDLCVRLFNOELDRPLREVMMAEPASVD 6542

```

```

QY 598 AALLDQTAFTQPALFTPEYALALMRSGVPELIVAGHSIGELVAAVCVAGVSLDEAVEL 657
DB 6543 SEELNQTCTTPALPANEVALFRLLBSWGLIPDVIYAGHSVEIILAAHVAGLSLEDAKL 6602
QY 658 VAAGRLMQALPAGGAMVSIAPADVAAVAAPHA--SVSIAVNPADQVVIAGAGQPVH 716
DB 6603 VSAALMQALPADGAMVAVQATEDE-----VLPHLDQVGIAAVNGPQSVVSGAEDVVV 6658
QY 717 AIAAAMAAKARITLHVSHPFSLAPMLAEAGRAVESVYRPSVLVSNISGKACT 776
DB 6659 AIGEFREGRKTRSLKVSHPFSLAPMLAEAGRAVESVYRPSVLVSNISGKACT 6717
QY 777 DEVSPGVWVHAREVVPFADGVKALAAAGAGTFVEVPGKSTLLGLVACMPDARPALA 836
DB 6718 -EPYTPYVWVHAREVVPFADGVKALAAAGAGTFVEVPGKSTLLGLVACMPDARPALA 6775
QY 837 SSRAGDEPATVLEALGGLMVAAGLVSVAGLPFGSGRRVPLPTYPMOERYWIDTKADA 896
DB 6776 AVRTDRPEPQAVVVALALHAGVSPDWRALFP-GARRVDLPYPFQYERLMLL-ABEE 6832
QY 897 ARGDRRAPGAGHDEVE--CGAV--RGDDRARSARD--HP-----PPES 934
DB 6833 YTGTAATGIGIGAHEPLAGAAVALPGAGGLVTGRLSTHTHMLADHTVMSVLLGCT 6892
QY 935 GRREKVEAGDRPRLTIDEPGVLDHLVLRTERRAPGLAGEVLAIDA-----AGLSFN 988
DB 6893 ALVELAVRAGEBAGCAQVEDLTLEAPLI--VPER--GVAVQVWVEBEEBGRCALSVH 6947
QY 989 D-----VOLALGMVDDLPGRKN-----PELLGSGCAGRIVAVGCVGLV 1030
DB 6948 SRREDPADPADVWVHVALGRLTDLALPEPHGADLGAMP--AGAAVLDLTFYEGLDIG 7004
QY 1031 VQO-PV-----IAL--SAGAPATH----- 1046
DB 7005 LGYGVFGRGLSVRSQGDVLAVALPEGTBAFAVHPALLDALHAAGGLVPLVDG 7064
QY 1047 -----YTTSALVLP----- 1056
DB 7065 PLLPAMSEVGRAGATALRVKLSRTCTDAVSLTVADSGGLVATGSLRPVSRQL 7124
QY 1057 -----RPOALSATEA-----AMP-VAVLTA-- 1076
DB 7125 LAAGSGRRADSLFGLAMQPAVLTETDPAAGLDVRSYADLALSTEBALPDVAVTVVP 7184
QY 1077 -----WYALDRIRAL--QPERVLIIAATGCVGLAAV 1106
DB 7185 CPAGSGRETADRVHSTSEVALVQVWMLAEERAGRLALVTRPGD--LAHAAVWGL--V 7238
QY 1107 OMAOH-----VGAE-----VAAT----- 1119
DB 7239 RASQSENDRILVLEAETDAVAVLPAALASGERPQAVRGEVFPRLVKAITSATGTA 7298
QY 1120 AGTPE-----KRAYLESIGRVYSDRSRFPV-----ADVRAW 1152
DB 7299 TGTDPFGVPLLTGASGALGLVAHRLVHAEGVRSLLLSRRGAEPGAVELBAELAAW 7358
QY 1153 -----TGGEQVUVLNSLS-----GELIDKSFVLLASHGRF----- 1183
DB 7359 GAEVWMAACDVADREAVASGMLNGEBSLSAVHTAGVIGDIAVSLTPRMRREVFRPKV 7418
QY 1184 -VELGRDCVADNOLGLRPL-----RMLSPSLV 1211
DB 7419 DAVLHNECTHD--MGLAAFYVPSVAGMVSAGQASVAAANSFLDANFSAHRRREGUPAI 7476
QY 1212 DLR-GMULERPARVALIE-ELGLIAGVFTPPPIATL-----PIARVA 1254
DB 7477 SLANGVWQSGAMTDGLVEADLRARMAASGVPLPEEBGLRFLFDALASDEAVLAPVRIDT 7536
QY 1255 DAFSMAQAQHLGLVLTLDPEVQIRIPTIAGAGPSGTGDDLDLDRLASAPABAALAE 1314
DB 7537 GALRAGEAPVPLRALV-----PAAARRTARAPAPAS--SLAREEMPEAEKEKTVL 7589
QY 1315 AFLRTQVQVL-RTPETKVAEALFTRLGMSDLMAVELRNRIEASLKSLSTFELSTSPN 1373

```

Db 7590 DLVFAEVAAGIGHASDRTPREHAFODLGFSDLSLAVELRNKLNKATGRLPATLVPDHP 7649
 Oy 1374 IALLAOLLDLALATA-----LSIERVAE---NLBAGVNDPVS 1410
 Db 7650 PALIAHVFMAAAGAEPAVDSLLADLDREOELVTTLAASEARDRLS 7699

RESULT 10

US-10-156-761-7963
 ; Sequence 7963, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156, 761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 7963
 ; LENGTH: 1635
 ; TYPE: PR1
 ; ORGANISM: Streptomyces avermectilis
 ; US-10-156-761-7963

Query Match 29.8%; Score 2152; DB 15; Length 1635;
 Best Local Similarity 34.1%; Pred. No. 3e-156;
 Matches 609; Conservative 185; Mismatches 554; Indels 436; Gaps 49;

Oy 4 RPIRAEDPDAIYACASCLPGVYIDLSGFWTLLEGSDTYGRVPAER-WMAAMPDPDP 62
 Db 25 RELHARQOEPIAIVGMSCRFPGVGSPPELWRLLAEGEDALSQFPVGRGMVDLYDPDP 84
 Oy 63 DAPKTPVTRASFSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENMAIAPSA 122
 Db 85 EOPKVTYTRCGGFHDAADPAEFEGISPREALMDPQRLLEAEMAEFAHAGIDPAGL 144
 Oy 123 VGTETGVFTIGIPSEY---EALPQATASAEIDAAGLGTMPVAGAGISYALGRPCV 179
 Db 145 KGSRTGVFAGLMYHDYFTPGVLPPEG---VEGLMSTGNGSVASGRITVTFEGEPVAV 199
 Oy 180 AVDTAVYSSSLVAVHQAOSLRSGECSTALAGVSLMISPTLVWLSTKTRALARPGRCA 239
 Db 200 TVDTACSSSLVLAHVAQALRTGBCSMALAGVTVMGPFAVEFSRGRGLAPGRCSF 259
 Oy 240 SAAEDGGRGEGCAVWVTKRLSGARADGDRLLAVIRGSAIHDGASGLTVPNGSSOEIV 299
 Db 260 AAAADGASWSRGTLLEVERLRDARRLGHKYLAVRGSAAVODASNGLTLPNPSOORV 319
 Oy 300 LKRALADAGCAASSVGVYEAHGTGTTGDPTEIGALNAVYAGRDVATPILLIGSVKTNL 359
 Db 320 ILLEALAGARLSAEEDVAEHAHTGTSIGDPTEAQAALATATGQERE--RPLLLGSLKSNIG 377
 Oy 360 HPEVASGTGLLKVYLSIHQOIPAHHLAQAOLNPRISGDLRLVYTRAKTPWPMNTPRR 419
 Db 378 HTQAAAGVAGVIAKVLAMQHVLPKTLHVPTRPVMSVGSVDLLTEALAMPETDHPRR 437
 Oy 420 AGVASSFGMSTNAHVLEAPATCTPPAPER-----PAELL-----VTSARTAS 464
 Db 438 AGVSSFGISGTHAIIIEQAPDEPAAGDPADDTAGGEGVAEELAKRTTVVWVLSAKKA 497
 Oy 465 ALDQAARLRDLHETYPSCIGDVAFSLATTRSAHEHRLAVAAATSRREGALDAAAQGO 524

Db 498 ALREQAQRLLSYVD-LAAASPADVGSLLATRSALREHRAAVGADGELRALITLAAAGE 556
 Oy 525 TSPCAVSIADSSRGKLAFLFTGCAQOTLGMGRGLVDMGAFAEAF-DLCVRLFNQELDR 583
 Db 557 PAAQVTVGRAGADRGKGFPLFGSGOSRIGMGELYAVYVFAAAYNEVCALL--DL-- 611
 Oy 584 PLREVMMAEPASVDALLDQTAFTOPALFTFEYALALMWSWGEVPELVAGHSIGELVAA 643
 Db 612 -----FVDVDAETLHQGTGSTOPALFAMEVALLFLLSMAIRPDVYVAGHSIGEIAAA 662
 Oy 644 CVAGVFLEDAVPLVAAARGLMQALPAGGMNLSIEAEADVAANAHAH-AASVSIANA 702
 Db 663 HVAQVSLIGDAARLVSRALMQALPAGGMVAVQAEDS---VLEYLTEGVEIAAING 718
 Oy 703 PDQVVIAGAGOPYAIAAAGARFKALVSHAFSPMAFMLBAFRAESVSRRP 762
 Db 719 POSVYVSGAEDAVAAVAEVAAGOKTSRLKVSAAFSPLMDPMLFEFAAVRGLTTEGP 778
 Oy 763 SYLVSNLSGKACTDEVSSPGYWRHAREVVRPADGVKALHAAGAGTFVEVGPSTLLGL 822
 Db 779 RIPVSNLTGRLA--EYTPHNVHVRHAEVRFADGVRTLHEIGVTFVEVGPGLVSAL 836
 Oy 823 VPACMPDAPALLAASRAGDEPATVLEALGLMAVGLVSNAGLFPSSGRVPLPTYPW 882
 Db 837 AOGCLDADIVTPALRTDRPEPOLTLVAGQLHTRGVSBDWOLF--GALRYDLPTYAF 895
 Oy 883 QRERYWI--DTKADDAARG-----DRAPG----- 905
 Db 896 QRERYWLKTTAVDVSSAGLSSGHALLGAVETAGQDVLFTGRSLPADLPMTAETTVG 955
 Oy 906 -----AGHDEV-----EGGAVR-----GGDRSA 925
 Db 956 GVPVLPVAAVELLASAGQAAGFPYLEQLDVDAVLAEGQAVRIVQIGADDDPDRRSV 1015
 Oy 926 RL-DHP-----PESGREKRYEA-----AG 944
 Db 1016 RVSRPRAAGSGSEBWRHVTGKLRNAPAEFDMIMPFGAEBGDEVRGLRAWRAG 1075
 Oy 945 DRPF---RLBIDE-----PGVLDHLVLRTERA-----PG 972
 Db 1076 DELYAELELPTERDRADRYALHPEVLDAALGMATSDRAADPGEVLMPHMAGSNVHTPG 1135
 Oy 973 LGEVEIAY-----DAAGLSFNDVQ-----LALGMVDD----- 1000
 Db 1136 PGAVRIRIRVGDVAVALIELSDADGSPVASVSRTRTMTAAQVLAARVAHQDMFEIDW 1195
 Oy 1001 -----LPGKPNPPL 1010
 Db 1196 VEHTAASAPATAVADVAVLDEGVHSDIDLARKAVDGAATPAHVLLPCAPSEGDAAA 1255
 Oy 1011 LGGECAGRIYAVGCVNG-----LVNGOPVIALSAGAF-ATHVYTTSA---ALVLRP 1058
 Db 1256 TRGLAGVLSVVOQVADERPSPDARLVMTRGAVELDTALDLAHVAVSGLVSAQTEHP 1315
 Oy 1059 QALSAIEA-----AAMPVAYLTAMYAL-----RIARLQPER- 1091
 Db 1316 DRILLVADDSVTNGTGTVDGVDPALAALALADEQAAVRODKLVVPHARITPPEEP 1375
 Oy 1092 -----VLHAATGVGLAAVQAQVGAE-----VHAATAGTPERKAYL 1129
 Db 1376 TGHAMDEGVTLTLGAAGGLGRSL--ARHIVARGARRLLLASRRPAAEGVGLVAAAL 1432
 Oy 1130 ESIG--VRVYSDRSR--FVADVRAWMTGGEGVDVNLSSGELLDKSFNLLRSHGRVEL 1186
 Db 1433 SELGAHVDAACDIADDALELASYPABEHPVAVVAAAALDDGVITALTP----- 1485
 Oy 1187 GKRDY---ADNQLGLRPLRLNLSFS---LVDLRGM-----LERP 1221
 Db 1486 GRUDTVLRPRADGALHHELTROLNLSAFVLFSSLAIVLSAGAAGVAAAANAFDGLARQ 1545
 Oy 1222 ARVRAL-----LEELGLIAAGVTPPP 1244
 Db 1546 RRAAGLPSTSLANGLMTDNGMGDRITDADRNRMGAGVVPFSPBAGLALFDALTVDRP 1605

QY 1215 GMLERPARVALLBEILGLIAGVFPPIATLPIARVADAPSMQAQHLGLVLT- 1272
Db 3292 A-----ERIQOMLEIMALFDAGVLLPLAMPVRBAHEALRVSOARVGVKVLTP 3345
QY 1273 -----LGDPEVOIRI- 1282
Db 3346 AALDAEGVLLTGAGTGLAVARHLVTEHDVRLLLVSRGVAPODLAAELGALAEVVA 3405
QY 1283 -----PTH-----AG----- 1287
Db 3406 ACDVANRKAALKALIEDIPPEHPVTGIVHTAGVLDGVVSGLTPERVDTLKPXDAALLTL 3465
QY 1288 -----AGPSTG----- 1293
Db 3466 ESVIGELDLDPALFYIESSAASMLGPGQGSYAAANOPLDTLARRARRGLTSVSLKGL 3525
QY 1294 -----DRD-----LDRRL----- 1301
Db 3526 WHEASGLTGLADIDRDMRAGIAPMPTDEALHFDRAATELGDPVLLPMRLNEALEDR 3585
QY 1302 -----ASAPAPAAALAEF-----LRTQV 1321
Db 3586 AADGTLPELISGLVRRHRPSARAGTATAPATGPEAFAREIAAADPRRALRDVLRGHV 3645
QY 1322 SQVL--RREIKVGAELFTRLGMSLMAVELRRIEASLKLKLTSTPLSTPNIALAQ 1379
Db 3646 ALVIGHSGPE--ALDEQAQRDIDGDSLTVAELRNKLNALNETGRLEPGLTFVFPYPPSALAD 3704
QY 1380 NLDALATA-----LSLEVAENIRAGVONDFVSSGADOD 1415
Db 3705 HLELLAPATQPTAPLAELEERV-EQLLSAASGCPASAVDEE 3748

RESULT 13
US-09-861-289-2
Sequence 2, Application US/09861289
Patent No. US20020110897A1
GENERAL INFORMATION:
APPLICANT: Sherman, D. H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600 438US1
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 5215
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-861-289-2

Query Match 29.5%; Score 2125; DB 10; Length 5215;
Best Local Similarity 37.0%; Pred. No. 1.7e-153;
Matches 581; Conservative 175; Mismatches 568; Indels 248; Gaps 46;

QY 1 VADRP---ERAA--EDPIAIVGASGRLPGVYIDLSGFWTLLEGSRDVTWGRVPAER-WDAA 55
Db 2812 VAERASGAGDRATDDPDAIVAGMACRYGVSSPEDMLRLVAEGTDAISEFPVNRGMWLE 2871
QY 56 AMFPDDPDAPGKTPVTRASFSLDVACPDASFSGISPREBALRMDPAHRLLEVCWEALENA 115
Db 2872 SLYPDPSKSTTYCRBEGFLEGAGDFDAFFGISPREALVMDPGORLLLEVMSEALERA 2931
QY 116 AIPASALVGTETGVTGIGPSEYEAALPOATASAEIDAHGGL--GTMPVGAGRISVLAG 173
Db 2932 GIDPSLSRSGGVVGAAGHSY--ASDPLVPEG---SEGYLITGSADAVMSGRISVALG 2987

QY 174 LRGCVAVDTAAYSSSLVAVHLACOSLSRSGECSTALAGVGLMTPSTLYWLSKTRALARD 233
Db 2988 LRGSMTEVETACSSSLVLAHLAVALRHGECGLALAGVAVMDPAFVFEFSRQKGLAAD 3047
QY 234 GRCKAFSAEADGFRGSGCAVVLKRLSGARADGDRILAVIRGSAINHDGASSGLTVPNG 293
Db 3048 GRCKAFSAADGTGMAGVGVLVLERLSDARRAGHTVGLVITGAVAVQDASNGLTAPNG 3107
QY 294 SSGEIVLKRLLAAGCAASSGVVEAHGCTTTGDPREIQALNAVVGGLGDVATPILLIGS 353
Db 3108 PAQQRVIAELADAGLSPEDVAEAGTGTGLDPTLEAGALLAASGRNSGDHPLWLG 3167
QY 354 VKTNLGRPEYASGITGLKTVLSTLQHQIAPHLHAQALNRPISGCDRLTVTRARTMPD 413
Db 3168 LKSNVIGAAAGVGVITKMLQALRHGLLPRTHADPPTHADSSGRVVLITSEVMQR 3227
QY 414 WNTPRRAGVSSFGKSGTNAHVLEADPAATCTPPAPF-----RPAP-----LVL 458
Db 3228 TGRPRRTGSAFGGCTNAHVLEAPF-----RPAPPAGEAPGSGRAAGABGPLAMV 3283
QY 459 SARTASALDAQARLRPHLETYSQCGLDVAFSLATRSMEHRLVAATSRGLRAALD 518
Db 3284 SGRDEPLRQARLRDLHLSRTPGARPDIATSLAATRAALPDHRAVLIGSDGELAALD 3343
QY 519 AAAGCOTSPGAVRSIADSSRGKLAFTTGGCAOTLGMGRGLYDWSAFREAFDLCVRLFN 578
Db 3344 ALAEGRGPAVAVGVRD--GRMAFLFTGGSGRAGAHLLHAHTFFAALDBVTDRLD 3402
QY 579 QEIDRLREVWMAEPASVDALLDQTAFTOPALFTFEYALAAIWRSGVEBELVAGHSIG 638
Db 3403 PLIGRPLGALLDARPGSPERAAALLDREYTOPALFAVAVLAHLLEHGMKRPDLILGHSVG 3462
QY 639 ELVAACVAGVPSLEDAFTVLAAGRLMOALPAGAMWSTIAPREADVAANAAPHAASIA 698
Db 3463 ELAAAHVAGVLDLDDACALVAARGRLMORLPFGAMWSVBAEGEDVALLAGBEDACVA 3522
QY 699 AVNAPDQVVIAGAGOPHAIAAAMAARGARTKALHVAHSPPLMAPLBAFGRAVSVS 758
Db 3523 AVNGPRSVVTSGAEEVAEAAQAAGRGRTTRLRVAHAHSPMLMDGMLGFRVNAAGLR 3582
QY 759 YRPSIVLNSLCK--ACTDEVSSPGYVVRHAREVRFADGVKALHAAGTFVEVGPKS 817
Db 3583 YRBEPLTVSVTVGRPARPELTGPDYVAQVEPVAFADAVTRAHLGARTFLETGPDG 3642
QY 818 TLGLVYACMPDARPALASSRAGRDEPAT-----VLEALGLMANVGLVSNAGLF 868
Db 3643 VLCMAEECLEDDTVALLPAIH---KPGTAPHGPAPALGALRAAAAAYGKARVDMAGMH 3698
QY 869 PSG-----GRVPLPTTYPMOREYWIIDTKADDAAGDRRAG-----AGHDEVEE 913
Db 3699 ADGPEGPARVELPVNAFRHRRVYL-----ATGRAADTDMMYRIGWDRLP- 3744
QY 914 GGAVRGDRRSAR--LDHPPE-----SGRREK--VEAAGRPFRLEIDEQVLDHLVLRV 965
Db 3745 --AVTGAARTAGRWLVTHPDSPRCRELSGHAERLARAAGASPVPLPVDAAPA--DRASFAA 3801
QY 966 TERAPGLG--EVEIADVAGISFNDVOLATGMVDDLPKGNPPLLIGGEAGRIVAV-- 1022
Db 3802 LIRSATPDRGTDAAPVAG-----LSL--LSEEDPHROHAPVPG-----VLATLS 3848
QY 1023 -----GEGVNGLV--VGOPIVA-----LSAGAPATHYTTSAVLVP----- 1056
Db 3849 LMQMEBEAVEARWCBSRAVAADDERPVGAGALMGGRVAAALRPFRRWGLVDLPA 3908
QY 1057 RPOA--LSAIEAAMP-----VAYLTAM--YALDRIR-----LOGSERVLIHAA 1097
Db 3909 SPGAHMAAVERIAGCEDDIAVAFSGSWGRLRFLPRDGGGRPAADAYPRRGVLTATGG 3968
QY 1098 TGGVGLAIVQACQVGA-----VHAATGPEK--- 1125
Db 3969 TGAIGGLARWLAAAGBHLALTSRRGPDAPGAAGLEABELLIGAKTTPAACDTADNGL 4028
QY 1126 -----RAYESLGVRYVS--DSRSDFVADVAM--TGEGGVVDVVLNLSGE 1168

Db 4029 ARVLRALPEDTPTLAVFHAAGVPOVTPILSRTPSEHFPADVAGKAAGAAHDELTRRELIGAG 4088
Qy 1169 LIDKSFNLSHGRFVELGKRDYADNQLGLRPFLLR----- 1204
Db 4089 L--DAFVLVSSGAGWGSAGGAYAAANALDALARRAADGLPATSIANGWGGGCGMA 4146
Qy 1205 -NLSFSLVDLRGMLEPAPVALLLELGLIAAGVTPPIAT-LPIARVADAFRSMAQ 1262
Db 4147 DEAGAEYLGRRGM--RPMAPVSALRAMATAIASG--EPCPTVHTDWERGEGFTAFRP 4201
Qy 1263 AOHILGKLVLLGDEVOIRLPTHAAG-----PSTGRDLIDRLASAPARAALAEFL 1317
Db 4202 S-----PIAGLGTTP-----GGRAAETPEEGNATAADLTALPAPALRTALRELY 4247
Qy 1318 RTOVSQVLRTPPE-1KVGAELFTRIGMDSLMAVELRNRIBASLTKLSTTFISTSPNIAL 1376
Db 4248 RARIALALGDDPRAVEAGEFRFPMGFDLSATVLRKGLASATGLDLPDLLLPDRTPLA 4307
Qy 1377 LAONLDALATA 1388
Db 4308 LAAHLELATA 4319

RESULT 14
US-09-860-846-2
; Sequence 2, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikomycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860.846
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105.537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-2

Query Match 29.5%; Score 2125; DB 10; Length 5215;
Best Local Similarity 37.0%; Pred. No. 1,7e-153;
Matches 581; Conservative 175; Mismatches 568; Indels 248; Gaps 46;

Qy 1 VADRPET---ERRA-EDRIATVGCSCRLPGVVIDLSGFTWLLSEGRDVTGKVPAR-WDA 55
Db 2812 VAEGASFGGRADDDDIATVAGMACRYPGVSSPEDIMRLVAEGTDIASEFPVNRGMDLE 2871
Qy 56 AWPDPDADAGKTPVTASFLSDVACPDASFGISPREALRMDPAHLLLEVCEALENA 115
Db 2872 SLVDPEDESGKTYCREGFLLEGAGDFDAFFGISPREALVMDQQLLEVSREALEERA 2931
Qy 116 AIPASALVETETGVFIGIGSEYEALPQATASAEIDAHGGL-GTWPVAGAGRISYALG 173
Db 2932 GIPPSIRGSRGGRGVYGAHGSY-ASDPRLVPEG---SEGVLTLGSADAVMSGRISYALG 2987
Qy 174 LRBPQAVDPATYASSIVAVHLLACOSLRSGESTLLAGVSLMSPSTLWMLSKTRALAD 233
Db 2988 LEEGSMVETVACSSSLVALHLAVALLRHEGGLLAGVAVMDPAFVEFSRCKGLAAD 3047
Qy 234 GRCKAFSAEADGFGRGSCAVVVKRLSGARADDRITIAVIRGSAIHHGASSGLTYPNG 293
Db 3048 GRCKAFSAADGTCMAGGVVTLERLSDARRAGHTVGLVTVGAVNODGASNGLTAPNG 3107
Qy 294 SSGEIVLKRALLADGCAASSVGVYEAHGTGTTGLDPIEIQALNNAVGLGSDVATPLLIGS 353

Db 3108 PAQORVIAELADAGSPEDVDVAEHAHTGTGRLGDPTEAGALLAASGRNSGDHPLWGS 3167
Qy 354 VKTNILGHEPYAGSGITGLKLVLLSQGOIPAHLLAQLNDRISMGDLRTVTRARTMPD 413
Db 3168 LKSNIGHAQAAGVGVITKQLRHLRGLLPRTHADEPTHADWSSGRVLLTSEVWQR 3227
Qy 414 WNTPRAGVSSFGMSGTNAVHLEAPPAATCTPPAPE-----RPAE-----LLVL 458
Db 3228 TGRPRRTGVSAFVGVTGNNAVHLEAPA-----PPAPEPAGEAGGSRAAGAEGLWV 3283
Qy 459 SARTASLMDQAARLDHLETYSQCLGDVAFSLATTRSMERLVAATSRBELRALD 518
Db 3284 SGRDEPALRQOARLDHLSRTGAPRDTAFSLAARAFDHRVAVLIGDGAELAAALD 3343
Qy 519 AAAGGTSFGAVSIDSRSKLAFLPTGGAGOTLIGRGGLYDWSAFREAFDLCVLFN 578
Db 3344 ALBGRGPAVAVGVRDRD-GRVAFLTGGSQAGAHDLHAAHTFPASALDEVTBRLD 3402
Qy 579 QELDRPLREYMAEPASVDALLDQTAFTQPALFTEYALAAALWRSVGEPELVAGHSIG 638
Db 3403 PLIGRPLGALLDARPGSPERALLDRTETQPALFAVEVALHRLLEHGMRPDLLGHSVG 3462
Qy 639 ELVAACVAGVFSLEDVFLVAARGRLMOALPAGCANVSTIAPEDVAAVAAPHAASVIA 698
Db 3463 ELAAAHVAGVLDIDDCALVAARGRLMORLPPOGANVSVAGDEVRALLAGREDAVCVA 3522
Qy 699 AVNAPQOVVLAGOPVHAIAAAMAARGARTKALHSHAHSPLMPLMEAFGRVAVSVS 758
Db 3523 AVNGPSSVSIAGEAVALAQAOLAGRGRTRLRVAAHSHSPMDGMLGFREVAAGLR 3582
Qy 759 YRRPSIVLSNLGK-ACTDEVSPGYWVHAAREVVRPADGVKLAHAAGTPEVVGPKS 817
Db 3583 YREBELTVSVTGRPARPELGTQPDYVWQVREPRFADAVTAHRLGARTFLETGPDG 3642
Qy 818 TLILGVPCMPDRPALASSRAGDEPAT-----VLEALGIMAVGLVSNAGLF 868
Db 3643 VLCGMAEECLEDDTVALLPAIH---KPGTAPHPGAPALRAAAYGARGADVWAGM 3698
Qy 869 PSG-----GRRVLPPTWQREERYWIDTKDADARGBRAG-----AGHDEVEE 913
Db 3699 ADGPBGPARRVELVHAFFHRRYWL-----AARGAADTDMMYRIGMDLPL- 3744
Qy 914 GGAVRGGRRSAR--LDHPPE-----SGRREK-VEAAGDRPRLIEDEPGVLDHLVLRV 965
Db 3745 --AVTGARTAGKWLVIHSPSPRCRELSGHAELALAAAGSVPLPVDADARA-DRASFAA 3801
Qy 966 TERRAPGLG-EVEIADVAGLSFNVOALGVPPDI-PGKPNPPLIGECAGRIYAV-- 1022
Db 3802 LIRSATGPDTRGDTAAPVAGV-----LSL-LSEEDRPHROHAPVPAG-----VILATLS 3848
Qy 1023 -----GEGVNGLV--VCGPVIA-----LSAGAPATHTTSAALVLP----- 1056
Db 3849 LMOAMEBEAVALKWCSSRAVAADDERPVGAAALMGIRVAALEPFRMGGLVDLPA 3908
Qy 1057 RPOA---LSAIEAAMP-----VAYLTAM-YALDRITAR-----LOPGERVLTHAA 1097
Db 3909 SPQAAHMAAVALHLAGBEDQIAVRASGWSGRRLTLRLPDGCGRTAARPAVPRGVLVTGG 3968
Qy 1098 TGGVGLAAVQMAOVGA-----VHAATGPEK-- 1125
Db 3969 TGAIGHLARWLAAAGAEHLALTSSRGPDAPGAAGLEAELLIGAKVTFAACTADRDGL 4028
Qy 1126 -----RAYLESIGVRYVS--DSRSRPFADVAVAM--TGEGVDVVLNLSGE 1168
Db 4029 ARVLRALPEDTPTLAVFHAAGVPOVTPILSRTPSEHFPADVAGKAAGAAHDELTRRELIGAG 4088
Qy 1169 LIDKSFNLSHGRFVELGKRDYADNQLGLRPFLLR----- 1204
Db 4089 L--DAFVLVSSGAGWGSAGGAYAAANALDALARRAADGLPATSIANGWGGGCGMA 4146
Qy 1205 -NLSFSLVDLRGMLEPAPVALLLELGLIAAGVTPPIAT-LPIARVADAFRSMAQ 1262
Db 4147 DEAGAEYLGRRGM--RPMAPVSALRAMATAIASG--EPCPTVHTDWERGEGFTAFRP 4201

1263 AOHGKLVLTGDEPEVOIRIPTHAGAG-----PSTGDRDLDRLASAPARAALAEFL 1317
Db 4202 S-----PLIAGIGTP-----GGGRAAETPEEGNATAADLTALPPEALRRLRLBELV 4247
Qy 1318 RTQVSQVLRTPE-IKVGAELFTRLGMDSLMAVELRNRIEASLKUKLSTFPLSTSPNIAL 1376
Db 4248 RARPAALGLDDPAEVAEGEFPFAMGFDLSLATVRLRRGLASATGLDLPDPLLFPDRTPAA 4307
Qy 1377 LAONLIDALATA 1388
Db 4308 LAHLAEILATA 4319

RESULT 15
US-09-988-384B-2
; Sequence 2, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin.
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-2

Query March 29.5%, Score 2125; DB 11; Length 5215;
Best Local Similarity 37.0%; Pred. No. 1.7e-153;
Matches 581; Conservative 175; Mismatches 568; Indels 248; Gaps 46;

Qy 1 VADRP1---ERRA-EDPIAIVGASCRLPGVYIDLSGFMTLEGSRDVTGRPAER-MDA 55
Db 2812 VAEGASGCDATDDPIAIVGACRYGVSSPBDMLRLVABGDALSEFPVRGMLE 2871
Qy 56 AWPDDDPAGKTPVTRASFSLSDVACFDASFFGISPREALRMDPAHRLLEVCWEALENA 115
Db 2872 SLVDPDESKCTTYCREGFLGAGDFDAFFGISPREALVMDPQORLLEVSWEALERA 2931
Qy 116 AIAESAVGTETGVFIGISPEVEALPQATASAEIDHGL--CTMPSVGAGRTSYALG 173
Db 2932 GIDSSLSGSGRYVGAHGSY-ASDRLVPEG--SEGYLLTGSADAVMSGRISYALG 2987
Qy 174 LRGCVAVDATYSSSLVAVHLACOSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARD 233
Db 2988 LEGSMYETETCSSSLVAVHLAVALRHGEGCLALAGVAAVMADPAAVEERSROKGLAAD 3047
Qy 234 GRCAFASEADDFGRGECACAVVILKRLSGAPADGDRILAVIRGSAINHDAASGGLTPNG 293
Db 3048 GRCAFAFSAADGTMAEGVGLVLERLSDARRAGHTVGLVGTAVVNDGASNGLTAPNG 3107
Qy 294 SSOSITVLKRALADGCAASSGVVEAHGTGTTIDPRIQALNVYGLGRVATPILIGS 353
Db 3108 PAQORVIAEALADGLSPEDDAVEAHGTGRLGDPLEAGALLAASGNRSGDHPMLIGS 3167
Qy 354 VKTNIHPEVASGTGLKLVLSLOHGOI PAHLHAQALNPRIISWGDLELTVTTRARTPPD 413
Db 3168 LKSNIGHQAALAGGVVTKMLQALRHGILLPRTLAHDEPTPAADSSGCVRLITSVPPQR 3227
Qy 414 WNTPRACVSSFGMSGTNAHVLEAPPAATCTPPAE-----RPAE-----LLVL 458
Db 3228 TGRBRRTGVSAFVGVTNAHVLEAPPA-----PPAPBPAGAPGGSRAAEGEGLAVVV 3283

Qy 459 SARTASALDAQARLRDHLLETYPSCGLDVAFSLATTRSAMERHLAATASTRREGIARAALD 518
Db 3284 SGRDEPRLRSQARRLRHLSTTPGARPRDIAFSLAATRAAFDRAVULIGSGAELAAALD 3343
Qy 519 AAGQOTSPGAVRSIADSRGKLAFLFTGGAQOTLGMRGLYDVWSAFREAFDLCVRLFN 578
Db 3344 ALAEGRGPAVVRGVRD-GRMAFLFTGQOSQAGAHDLHAHNTFFASALDEVTRLD 3402
Qy 579 QELDPRFVWMAEPASVDALLDQTAFTQPALFTFEYALALMRSGVPELVAGHSIG 638
Db 3403 PLIGRPLGALLDARPGSPEALDRTETOPALFAVEVALHRLLEHGMKPPDLLGLHSVG 3462
Qy 639 ELVACVAGVFSLEDAVFLVAARGLMQALPAGGAMVSIEPEADVAVAAPHAASVIA 698
Db 3463 ELAAHVAAGVLDLDDCALVAARGRLQRLPFGAMVSVAAGEVEVALLAGREDACVA 3522
Qy 699 AVNAPDOVVIAGAPVHAIAAAMAAAGARTKALHVAHSPPLMAEMLEAFGRVAESVS 758
Db 3523 AVNGPRSVVTSAGEAFAAQAQLAGRGRTRLRVAAHAFSPPLMDGMLAGFREVAAGLR 3582
Qy 759 YRPSIVLVNSLCK-ACTDEVSPGYWYHAREVVFADGVKALHAAGCTFVEVPPKS 817
Db 3583 YREBELTVVSTVGRPARPGBELTGPDYVAQVREPVAFADAVRTAHLGARTFLETGPDG 3642
Qy 818 TLGLVPAWMDARPALASSRAGDEPAT-----VLEALGLMAVAGLVSMAGLF 868
Db 3643 VLCMAEECELDQVALLPAIH---KPTAPHPGPAAPGALRAAANAAYGCAARDMAAGM 3688
Qy 869 PSG---GRVPLFTYPMQERYWIDTKADDAAGDRARQ-----AGHDEVEE 913
Db 3699 ADGPEPARVELPVHAFFRRHVVLT-----AGRAADTDMMYRICMDRLP- 3744
Qy 914 GGAVRGDRSAR--LDHPPE-----SGRREK-VEAAGRPFLLEIDEGVLDHLVLRV 965
Db 3745 --AVTGARTAGRLVTHPDSPRCRLSGHAERLRAAGSPVLPVADAPA--DRASFAA 3801
Qy 966 TERAPGLG-EVEIAYVDAAGLSFNDVQALGMVDDLPGRKNPPLIGECACGRIVAV-- 1022
Db 3802 LIRSATPDRTRGDTAAVAGV-----LSL-LSEDEPHQOHAPVPAAG-----VLATLS 3848
Qy 1023 ---GEGVNGLV--VGOPYIA-----LSAGAFTHVTTSAVLVP----- 1056
Db 3849 IMQAMEEAAVEARVWCVSRAAADAERPRVGAALMGRAVALRPRFMGGLVLDPA 3908
Qy 1057 RPOA---LSAIEAAMP-----VAYLTAW-YALDRIR-----LQGEVVLHAA 1097
Db 3909 SPGAHMAAAVERLAGEDQIAVRASSGWRRLTRLRDPGGGRTAADAYRPRGTVLVTGG 3968
Qy 1098 TGGVGLAAVQMAQIVGAE-----VHATAGTPEK--- 1125
Db 3969 TGLGHLARWLAAGAHEHLATLSRRGPDPAAGAAGLEALLGAKTTPAACOTADNDGL 4028
Qy 1126 -----RAYLESIGVRYS--DSRSDRFVADVRAM--TGEGVDVVLNLSGE 1168
Db 4029 ARVLRAIRBEDPTPLVAFAHAGVPQVTPLSRTPSEHFAVYVAGKAAGAHNDLRELGAG 4088
Qy 1169 LIDKSFMLLSHGKRFVELGRDCYADNOLGIRPLR----- 1204
Db 4089 L--DAFVLSSGAGVWSSAGQAAANAALDALARRAADGLPATSIAWGWSGGMGGA 4146
Qy 1205 -NLSFSLVDIRGMMLEPARVALLEELGLIAGVTPPIAT-LPIARVADAFRMAQ 1262
Db 4147 DECAEYVIGRGM---RPMAPVSLRMAATRIASG--EPCPTVTHTWMEFGEFTAFRP 4201
Qy 1263 AOHGKLVLTGDEPEVOIRIPTHAGAG-----PSTGDRDLDRLASAPARAALAEFL 1317
Db 4202 S-----PLIAGIGTP-----GGGRAAETPEEGNATAADLTALPPEALRRLRLBELV 4247
Qy 1318 RTQVSQVLRTPE-IKVGAELFTRLGMDSLMAVELRNRIEASLKUKLSTFPLSTSPNIAL 1376
Db 4248 RARPAALGLDDPAEVAEGEFPFAMGFDLSLATVRLRRGLASATGLDLPDPLLFPDRTPAA 4307

Sat Oct 4 14:14:29 2003

us-10-014-717-2.rapb

Page 17

Oy	1377	LAQNLDA	1388
		:	
Db	4308	LAHLELATA	4319

Search completed: October 2, 2003, 17:47:41
Job time : 109.09 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:33 ; Search time 34.7527 Seconds
(without alignments)
1730.044 Million cell updates/sec

Title: US-10-014-717-2
Perfect score: 7210
Sequence: 1 VADRPTEAEDPPIAIVGAS.....GVQNDPVSSGADQDWETIAL 1421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA: *
1: /cgn2_6/ptodata/1/iaa/5A COMB.pdp: *
2: /cgn2_6/ptodata/1/iaa/5B COMB.pdp: *
3: /cgn2_6/ptodata/1/iaa/6A COMB.pdp: *
4: /cgn2_6/ptodata/1/iaa/6B COMB.pdp: *
5: /cgn2_6/ptodata/1/iaa/PTCUS COMB.pdp: *
6: /cgn2_6/ptodata/1/iaa/Backfile1.pdp: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7210	100.0	1421	3 US-09-335-409-2	Sequence 2, Appli
2	7210	100.0	1421	4 US-09-568-102-2	Sequence 2, Appli
3	7210	100.0	1421	4 US-09-567-969-2	Sequence 2, Appli
4	7210	100.0	1421	4 US-09-568-480-2	Sequence 2, Appli
5	7210	100.0	1421	4 US-09-568-486-2	Sequence 2, Appli
6	7210	100.0	1421	4 US-09-568-472-2	Sequence 2, Appli
7	7210	100.0	1421	4 US-09-567-899-2	Sequence 2, Appli
8	4137	57.4	2439	3 US-09-335-409-7	Sequence 7, Appli
9	4137	57.4	2439	4 US-09-568-102-7	Sequence 7, Appli
10	4137	57.4	2439	4 US-09-567-969-7	Sequence 7, Appli
11	4137	57.4	2439	4 US-09-568-480-7	Sequence 7, Appli
12	4137	57.4	2439	4 US-09-568-486-7	Sequence 7, Appli
13	4137	57.4	2439	4 US-09-568-472-7	Sequence 7, Appli
14	4137	57.4	2439	4 US-09-567-899-7	Sequence 7, Appli
15	3831.5	53.1	2259	3 US-09-413-814-70	Sequence 70, Appli
16	3311.5	45.9	7257	3 US-09-335-409-5	Sequence 5, Appli
17	3311.5	45.9	7257	4 US-09-568-102-5	Sequence 5, Appli
18	3311.5	45.9	7257	4 US-09-567-969-5	Sequence 5, Appli
19	3311.5	45.9	7257	4 US-09-568-480-5	Sequence 5, Appli
20	3311.5	45.9	7257	4 US-09-568-486-5	Sequence 5, Appli
21	3311.5	45.9	7257	4 US-09-568-472-5	Sequence 5, Appli
22	3311.5	45.9	7257	4 US-09-567-899-5	Sequence 5, Appli
23	2273	31.5	2152	3 US-09-036-987A-3	Sequence 3, Appli
24	2273	31.5	2152	3 US-09-370-700-3	Sequence 3, Appli
25	2273	31.5	2152	4 US-09-603-207-3	Sequence 3, Appli
26	2250	31.2	6095	3 US-09-144-085-2	Sequence 2, Appli
27	2245.5	31.1	2890	3 US-09-413-814-67	Sequence 67, Appli

28	2245.5	31.1	3798	3 US-09-335-409-6	Sequence 6, Appli
29	2245.5	31.1	3798	4 US-09-568-102-6	Sequence 6, Appli
30	2245.5	31.1	3798	4 US-09-567-969-6	Sequence 6, Appli
31	2245.5	31.1	3798	4 US-09-568-480-6	Sequence 6, Appli
32	2245.5	31.1	3798	4 US-09-568-486-6	Sequence 6, Appli
33	2245.5	31.1	3798	4 US-09-568-472-6	Sequence 6, Appli
34	2245.5	31.1	3798	4 US-09-567-899-6	Sequence 6, Appli
35	2194	30.4	5087	3 US-09-144-085-1	Sequence 1, Appli
36	2150	29.8	1841	2 US-08-804-227C-6	Sequence 6, Appli
37	2128	29.5	3816	3 US-09-428-517-3	Sequence 3, Appli
38	2127	29.5	3170	3 US-09-036-987A-4	Sequence 4, Appli
39	2127	29.5	3170	3 US-09-370-700-4	Sequence 4, Appli
40	2127	29.5	3170	3 US-09-603-207-4	Sequence 4, Appli
41	2125	29.5	5215	3 US-09-105-537-2	Sequence 2, Appli
42	2113	29.3	4928	3 US-09-036-987A-5	Sequence 5, Appli
43	2113	29.3	4928	3 US-09-370-700-5	Sequence 5, Appli
44	2113	29.3	4928	3 US-09-603-207-5	Sequence 5, Appli
45	2108.5	29.2	3562	4 US-09-679-279-14	Sequence 14, Appli

ALIGNMENTS

RESULT 1					
US-09-335-409-2					
Sequence 2, Application US/09335409					
Patent No. 6121029					
GENERAL INFORMATION:					
APPLICANT: Schnupp, Thomas					
APPLICANT: Lignon, James					
APPLICANT: Molnar, Istvan					
APPLICANT: Zitzke, Ross					
APPLICANT: Cyr, Devon					
APPLICANT: Goerlach, Joern					
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES					
FILE REFERENCE: 4-30582A					
CURRENT APPLICATION NUMBER: US/09/335,409					
CURRENT FILING DATE: 1999-06-17					
NUMBER OF SEQ ID NOS: 30					
SOFTWARE: Patentin Ver. 2.0					
SEQ ID NO 2					
LENGTH: 1421					
TYPE: PRT					
ORGANISM: Sorangium cellulosum					
US-09-335-409-2					
Query Match					
Best Local Similarity 100.0%; Score 7210; DB 3; Length 1421;					
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	VADRPTEAEDPPIAIVGASCRIPGVYDLSGFWTLLGSRDPTVGRVPAERWDAAMFDP	60		
DB	1	VADRPTEAEDPPIAIVGASCRIPGVYDLSGFWTLLGSRDPTVGRVPAERWDAAMFDP	60		
QY	61	DDPAFGKTPYTRASFSDVACPDASFPGISPREALRMDPAHRLLEVCWEALENAAIAPS	120		
DB	61	DDPAFGKTPYTRASFSDVACPDASFPGISPREALRMDPAHRLLEVCWEALENAAIAPS	120		
QY	121	ALVETGTVFVIGIPSEYEALPQATASAEIDAHGIGCTMPVSGAGIISTALGIRGCVA	180		
DB	121	ALVETGTVFVIGIPSEYEALPQATASAEIDAHGIGCTMPVSGAGIISTALGIRGCVA	180		
QY	181	VDPAYSSLVAVHACQSLRSGRSTLACGVSIMLSPSTLVMLSKTRALARDRCRAFS	240		
DB	181	VDPAYSSLVAVHACQSLRSGRSTLACGVSIMLSPSTLVMLSKTRALARDRCRAFS	240		
QY	241	AEADGFRGSGCAVVLKRLSGARADDRILAVIRGSAINHDSAGSLTVPNGSSQEIIVL	300		
DB	241	AEADGFRGSGCAVVLKRLSGARADDRILAVIRGSAINHDSAGSLTVPNGSSQEIIVL	300		
QY	301	KRALADGCAASSGVYEAGTGTLLDPTIEIQALNNAVGLGRVATPPLIGSVKTNLGH	360		
DB	301	KRALADGCAASSGVYEAGTGTLLDPTIEIQALNNAVGLGRVATPPLIGSVKTNLGH	360		

361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRTYTRARTPMPDMNTPPRA 420
361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRTYTRARTPMPDMNTPPRA 420
421 GVSSFGMSGTNAHVLEBPAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480
421 GVSSFGMSGTNAHVLEBPAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480
481 PSOCIGDVAFSLATTRSAMEHRLAVALTSRGLRALDAAAOGQTSFGAVSISADSSRGK 540
481 PSOCIGDVAFSLATTRSAMEHRLAVALTSRGLRALDAAAOGQTSFGAVSISADSSRGK 540
541 LAFLEFTGGAGOTLGMGRGLYDWSAFREAFDLCYRLFNOELDRPLREVMMAEPASVDAL 600
541 LAFLEFTGGAGOTLGMGRGLYDWSAFREAFDLCYRLFNOELDRPLREVMMAEPASVDAL 600
601 LDQTAFTOPALFTFEYALALMRSGVEPELVAGHSIGELVAAQVAGVFSLEDAVFLVAA 660
601 LDQTAFTOPALFTFEYALALMRSGVEPELVAGHSIGELVAAQVAGVFSLEDAVFLVAA 660
661 RGRLEMOALPAGGAMVSIAPADVAAPAAVAPAAVSIAPVAPADVVIAGAGQPVHAI 720
661 RGRLEMOALPAGGAMVSIAPADVAAPAAVAPAAVSIAPVAPADVVIAGAGQPVHAI 720
721 AMAARGATKALHVAHSHAPMLAFAFGVAVSVSRRSIVLVSNTLSKACTDVS 780
721 AMAARGATKALHVAHSHAPMLAFAFGVAVSVSRRSIVLVSNTLSKACTDVS 780
781 SPGYVVRHAREVVRADGVKALHAAGAGTFVEVGEKSTLLGLVPACMPDARALASSRA 840
781 SPGYVVRHAREVVRADGVKALHAAGAGTFVEVGEKSTLLGLVPACMPDARALASSRA 840
841 GRDEPATVLEALGSLMAVGLVSMAGLFPSSGGRVPLPTYPWQERBYMDTYADDAAGD 900
841 GRDEPATVLEALGSLMAVGLVSMAGLFPSSGGRVPLPTYPWQERBYMDTYADDAAGD 900
901 RRAAPAGHDEVEEGAVGGRDRSARLDHPPESGRREVEAAGRPRLTIDEQVUDH 960
901 RRAAPAGHDEVEEGAVGGRDRSARLDHPPESGRREVEAAGRPRLTIDEQVUDH 960
961 LVLRATERRAPGLGEVEIAVDAAGSFPNDVOLAMVPDDLPCKENPPLALGCECAGRI 1020
961 LVLRATERRAPGLGEVEIAVDAAGSFPNDVOLAMVPDDLPCKENPPLALGCECAGRI 1020
1021 AVGECVNGVLVGOPIVIALSAGAFATHVTTSAALVLPPOALSIAIPAAAMPVAYLTAM 1080
1021 AVGECVNGVLVGOPIVIALSAGAFATHVTTSAALVLPPOALSIAIPAAAMPVAYLTAM 1080
1081 DRIARLOGERVLIHAATGVGLAIVOMAQHVGAHVHATAGTPEKRAVLESIGVAVSDS 1140
1081 DRIARLOGERVLIHAATGVGLAIVOMAQHVGAHVHATAGTPEKRAVLESIGVAVSDS 1140
1141 RSDREAVADVAMTGEVGVVNVLSIGELIDKSFNLSHGRFVLEGRKDCYADQOLGR 1200
1141 RSDREAVADVAMTGEVGVVNVLSIGELIDKSFNLSHGRFVLEGRKDCYADQOLGR 1200
1201 PFLRNLSFSLVDLRGMLEPRPARVALLLELLGLIAAGVTPPIATLPIARVADAFRSM 1260
1201 PFLRNLSFSLVDLRGMLEPRPARVALLLELLGLIAAGVTPPIATLPIARVADAFRSM 1260
1261 AOAQHLGLVLTLDGPEVOIRIPTHAGAGSTGDBDLDRLASAPAPAAAALAEFLRTO 1320
1261 AOAQHLGLVLTLDGPEVOIRIPTHAGAGSTGDBDLDRLASAPAPAAAALAEFLRTO 1320
1321 VSQVARTPEIKVGAALFTRLGMDSIMAVELNRIEASIKLSTTFSTSPNIALAON 1380
1321 VSQVARTPEIKVGAALFTRLGMDSIMAVELNRIEASIKLSTTFSTSPNIALAON 1380
1381 LLDALATALSLERVAENIRAGVQNDPVSSGADQDWEIAT 1421
1381 LLDALATALSLERVAENIRAGVQNDPVSSGADQDWEIAT 1421

RESULT 2
US-09-568-102-2
; Sequence 2, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1421, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

1 VADRPBIAADPRAIVGASCRLLPGVYDLSGFMTLLGSRDVTGRVPAERMDAAAFDP 60
1 VADRPBIAADPRAIVGASCRLLPGVYDLSGFMTLLGSRDVTGRVPAERMDAAAFDP 60
61 DPDAFGKTPVTRASFSLDVACFDASFFGISPREALRMDPAHLLLEVCEALENNAIAPS 120
61 DPDAFGKTPVTRASFSLDVACFDASFFGISPREALRMDPAHLLLEVCEALENNAIAPS 120
121 ALVGTETGVPIGIGBSYEALPQATASAEIDAAGGCTMPSVQAGRISTALGRPCVA 180
121 ALVGTETGVPIGIGBSYEALPQATASAEIDAAGGCTMPSVQAGRISTALGRPCVA 180
181 VDTAYSSSLVAVHLACOSLRSGECSTALAGVSLMSPSTLVMLSKTRALARDGRKAFS 240
181 VDTAYSSSLVAVHLACOSLRSGECSTALAGVSLMSPSTLVMLSKTRALARDGRKAFS 240
241 AEADGFGRCACAVVVLKRLSGARADGDRILAVIRGSAINHAGASSGLTVPNGSSQETVL 300
241 AEADGFGRCACAVVVLKRLSGARADGDRILAVIRGSAINHAGASSGLTVPNGSSQETVL 300
301 KRALLADAGCAASSVGYEVAHGTGTTLDPIEIOALNMYGUGRVATPPLIGSVKTNLGH 360
301 KRALLADAGCAASSVGYEVAHGTGTTLDPIEIOALNMYGUGRVATPPLIGSVKTNLGH 360
361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRTYTRARTPMPDMNTPPRA 420
361 PEYASGITGLKVVLSLOHGOI PAHLHAQALNPRISMGDLRTYTRARTPMPDMNTPPRA 420
421 GVSSFGMSGTNAHVLEBPAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480
421 GVSSFGMSGTNAHVLEBPAPATCTPPAPERPAELVLSARTASALDQAARLBDHLETY 480
481 PSOCIGDVAFSLATTRSAMEHRLAVALTSRGLRALDAAAOGQTSFGAVSISADSSRGK 540
481 PSOCIGDVAFSLATTRSAMEHRLAVALTSRGLRALDAAAOGQTSFGAVSISADSSRGK 540
541 LAFLEFTGGAGOTLGMGRGLYDWSAFREAFDLCYRLFNOELDRPLREVMMAEPASVDAL 600
541 LAFLEFTGGAGOTLGMGRGLYDWSAFREAFDLCYRLFNOELDRPLREVMMAEPASVDAL 600
601 LDQTAFTOPALFTFEYALALMRSGVEPELVAGHSIGELVAAQVAGVFSLEDAVFLVAA 660
601 LDQTAFTOPALFTFEYALALMRSGVEPELVAGHSIGELVAAQVAGVFSLEDAVFLVAA 660

QY 661 RGRIMQALPAGANVSIIEAPADVAAPAAVAPPAASVSIANAAPQVVIAGAGQEVHATAA 720
DB 661 RGRIMQALPAGANVSIIEAPADVAAPAAVAPPAASVSIANAAPQVVIAGAGQEVHATAA 720
QY 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
DB 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
QY 781 SPGWVRRHAREVVPADGVKALHAAAGACTFVEVGGKSTLLGLVPCMPDAPALLASSRA 840
DB 781 SPGWVRRHAREVVPADGVKALHAAAGACTFVEVGGKSTLLGLVPCMPDAPALLASSRA 840
QY 841 GRDEPATVLEALGLMAVGLVSWAGLPSSGRRVPLPTYPWQERWIDTKADDAARGD 900
DB 841 GRDEPATVLEALGLMAVGLVSWAGLPSSGRRVPLPTYPWQERWIDTKADDAARGD 900
QY 901 RRAAGAGHDEVEEGAVRGDRRSARLDHPPESGREKEVAAGDRPRLIEDPGVLH 960
DB 901 RRAAGAGHDEVEEGAVRGDRRSARLDHPPESGREKEVAAGDRPRLIEDPGVLH 960
QY 961 LVLETERBARGLBEVEIAVDAAGLSFMDVOLALGMVDDLLPKXNPPLLGGEGAGRIY 1020
DB 961 LVLETERBARGLBEVEIAVDAAGLSFMDVOLALGMVDDLLPKXNPPLLGGEGAGRIY 1020
QY 1021 AVGGVNGLVVGGOPVIALSAGAFTHVTSAAVLPRPOLSAIEAAMPVAVILTAWYAL 1080
DB 1021 AVGGVNGLVVGGOPVIALSAGAFTHVTSAAVLPRPOLSAIEAAMPVAVILTAWYAL 1080
QY 1081 DRIARLPGEREVLIIHAATGVGLAAVQWAQVGAEVHATAGTPEKRAYLESIGVRYVSDS 1140
DB 1081 DRIARLPGEREVLIIHAATGVGLAAVQWAQVGAEVHATAGTPEKRAYLESIGVRYVSDS 1140
QY 1141 RSDRFVADVAVRWATGEGVDVVLNSISGELIDKSPILARSHGFPVELGKRDCAONOLGR 1200
DB 1141 RSDRFVADVAVRWATGEGVDVVLNSISGELIDKSPILARSHGFPVELGKRDCAONOLGR 1200
QY 1201 PFLNLSFSVLVDLKGMMLEPARVVALLEELIGLIAAGVFTPPPIATLPIARVADAFRSM 1260
DB 1201 PFLNLSFSVLVDLKGMMLEPARVVALLEELIGLIAAGVFTPPPIATLPIARVADAFRSM 1260
QY 1261 AQAHLGKLVLTICDPEVOIRIPTHAGAPSTGDRDILLRLASAPARAALAEAFRLTQ 1320
DB 1261 AQAHLGKLVLTICDPEVOIRIPTHAGAPSTGDRDILLRLASAPARAALAEAFRLTQ 1320
QY 1321 VSQVLRPEIKVGAELFTRLGMSLMAVELRNRIEASLKLTSTFLSTSPNTALLAQN 1380
DB 1321 VSQVLRPEIKVGAELFTRLGMSLMAVELRNRIEASLKLTSTFLSTSPNTALLAQN 1380
QY 1381 LLDALATLALSLERVAEENLRAGVQNDFFVSGADODMEITIAL 1421
DB 1381 LLDALATLALSLERVAEENLRAGVQNDFFVSGADODMEITIAL 1421

RESULT 3
US-09-567-969-2
; Sequence 2, Application US/09567969
; Patent No. 635457
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zikler, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,969
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 2
; LENGTH: 1421
; TYPE: PR
; ORGANISM: Sorangium cellulosum
US-09-567-969-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRPERRAADPPIAIVGASGRLPFGVITDLSGFTLLBGSUDTGRVPAERMDAAWADP 60
DB 1 VADRPERRAADPPIAIVGASGRLPFGVITDLSGFTLLBGSUDTGRVPAERMDAAWADP 60
QY 61 DPDAFGKTPVTRASFSLDVACFDSPFGISPREALRMDPARHLLBVCHEALENAATAPS 120
DB 61 DPDAFGKTPVTRASFSLDVACFDSPFGISPREALRMDPARHLLBVCHEALENAATAPS 120
QY 121 ALVGTETGVFIGIPSEYEALPQATASAEIDAHGGLGTMPSVAGRISVALGRPCVA 180
DB 121 ALVGTETGVFIGIPSEYEALPQATASAEIDAHGGLGTMPSVAGRISVALGRPCVA 180
QY 181 VDTAYSSSLVAVHLACQSLNSGECSTALAGVSLMSPSTLVMLSKTRALARDCRKAFS 240
DB 181 VDTAYSSSLVAVHLACQSLNSGECSTALAGVSLMSPSTLVMLSKTRALARDCRKAFS 240
QY 241 AADGFGRGECACAVVYLKRLSGARADDDRILAVIRGSAINDGASSGLTVPNGSSQETVL 300
DB 241 AADGFGRGECACAVVYLKRLSGARADDDRILAVIRGSAINDGASSGLTVPNGSSQETVL 300
QY 301 KRALADGCAASVGYEBAHGTGTTLDPIETIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
DB 301 KRALADGCAASVGYEBAHGTGTTLDPIETIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
QY 361 PEVAGSITGLIKVLSIQHQIIPALHQAOLNPRISGDLRLVYTRARTPMDNTPERRA 420
DB 361 PEVAGSITGLIKVLSIQHQIIPALHQAOLNPRISGDLRLVYTRARTPMDNTPERRA 420
QY 421 GVSSFGMSGNAHVLEAPAACTPPAPERPALVLSARTASALDAQARLDHLETY 480
DB 421 GVSSFGMSGNAHVLEAPAACTPPAPERPALVLSARTASALDAQARLDHLETY 480
QY 481 PSQCLGVAFSLATTRSAMEHRLVAATSREGLRALDAAQOGTSPGAVSISDSRGK 540
DB 481 PSQCLGVAFSLATTRSAMEHRLVAATSREGLRALDAAQOGTSPGAVSISDSRGK 540
QY 541 LAFLEFGQAGTIGWGRGLYDWSAFREBAPDLVRLFNQELDRPLREVMMAEPASVDAAL 600
DB 541 LAFLEFGQAGTIGWGRGLYDWSAFREBAPDLVRLFNQELDRPLREVMMAEPASVDAAL 600
QY 601 LDQTAFTQPALFTFEVYLAALMRSGVPELVAGHSIGELVAAACVAGVFSLEDAVFLVA 660
DB 601 LDQTAFTQPALFTFEVYLAALMRSGVPELVAGHSIGELVAAACVAGVFSLEDAVFLVA 660
QY 661 RGRIMQALPAGANVSIIEAPADVAAPAAVAPPAASVSIANAAPQVVIAGAGQEVHATAA 720
DB 661 RGRIMQALPAGANVSIIEAPADVAAPAAVAPPAASVSIANAAPQVVIAGAGQEVHATAA 720
QY 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
DB 721 AMAARGARTKALHVSASHFSPMLAEMLEAFGRVAESVYRRPSIVLSNLSGKACTDEVS 780
QY 781 SPGWVRRHAREVVPADGVKALHAAAGACTFVEVGGKSTLLGLVPCMPDAPALLASSRA 840
DB 781 SPGWVRRHAREVVPADGVKALHAAAGACTFVEVGGKSTLLGLVPCMPDAPALLASSRA 840
QY 841 GRDEPATVLEALGLMAVGLVSWAGLPSSGRRVPLPTYPWQERWIDTKADDAARGD 900
DB 841 GRDEPATVLEALGLMAVGLVSWAGLPSSGRRVPLPTYPWQERWIDTKADDAARGD 900
QY 901 RRAAGAGHDEVEEGAVRGDRRSARLDHPPESGREKEVAAGDRPRLIEDPGVLH 960
DB 901 RRAAGAGHDEVEEGAVRGDRRSARLDHPPESGREKEVAAGDRPRLIEDPGVLH 960

QY 961 LVLRTERARRPGLGEVEIAVDAAGISFNDVOLALGMVDDLPKKNPPLLGGECAGRIY 1020
 DB 961 LVLRTERARRPGLGEVEIAVDAAGISFNDVOLALGMVDDLPKKNPPLLGGECAGRIY 1020
 QY 1021 AVGEVNGLVVGGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080
 DB 1021 AVGEVNGLVVGGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080
 QY 1081 DRIARLQGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAVLESIGVYSDS 1140
 DB 1081 DRIARLQGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAVLESIGVYSDS 1140
 QY 1141 RSDRVAADVRAVTGEGVDVNLNSGELIDKSFNLSHGRFVELGRDCCADNQLGR 1200
 DB 1141 RSDRVAADVRAVTGEGVDVNLNSGELIDKSFNLSHGRFVELGRDCCADNQLGR 1200
 QY 1201 PFLRLSFLVDLRGMLEPRPARVRLLELLGLIAAGVFTPPPIATLPPIARVADAFRSM 1260
 DB 1201 PFLRLSFLVDLRGMLEPRPARVRLLELLGLIAAGVFTPPPIATLPPIARVADAFRSM 1260
 QY 1261 AQAQHLGKLVLTLDPEVOIRIPTHAGAPSTGDRDLDRLASAAPARAALBAFLRQ 1320
 DB 1261 AQAQHLGKLVLTLDPEVOIRIPTHAGAPSTGDRDLDRLASAAPARAALBAFLRQ 1320
 QY 1321 VSQVTRTEIKKGAALFTRLGMDSLMAVELNRIEASIKLSTTFSTFNIALAON 1380
 DB 1321 VSQVTRTEIKKGAALFTRLGMDSLMAVELNRIEASIKLSTTFSTFNIALAON 1380
 QY 1381 LLDALATATLSLERAENLRAGVQNDPVSSGADQDMETIAL 1421
 DB 1381 LLDALATATLSLERAENLRAGVQNDPVSSGADQDMETIAL 1421

RESULT 4
 US-09-568-480-2
 Sequence 2, Application US/09568480

Patent No. 6355458
 GENERAL INFORMATION:
 APPLICANT: Schupp, Thomas
 APPLICANT: Ligon, James
 APPLICANT: Molnar, Istvan
 APPLICANT: Zinkle, Ross
 APPLICANT: Cyt. Devco
 APPLICANT: Goetlich, Joern
 TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
 FILE REFERENCE: 4-30582A
 CURRENT APPLICATION NUMBER: US/09/568,480
 CURRENT FILING DATE: 2000-05-10
 PRIOR APPLICATION NUMBER: 09/335,409
 PRIOR FILING DATE: 1999-06-17
 NUMBER OF SEQ ID NOS: 30
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 2
 LENGTH: 1421
 TYPE: PRT
 ORGANISM: Sorangium cellulosum
 US-09-568-480-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRIEBAEDPIAIVGASCRLLPGVVIDLSGFMTLLBGSRTDVGVRPAERMDAAWDP 60
 DB 1 VADRIEBAEDPIAIVGASCRLLPGVVIDLSGFMTLLBGSRTDVGVRPAERMDAAWDP 60
 QY 61 DPDAKGTPTWRASLTVACFDASFFGISPREARLMDPAHRLLEVCHEALENAAIAP 120
 DB 61 DPDAKGTPTWRASLTVACFDASFFGISPREARLMDPAHRLLEVCHEALENAAIAP 120
 QY 121 ALVGTETGVFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGRCPCA 180
 DB 121 ALVGTETGVFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGRCPCA 180

DB 121 ALVGTETGVFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGRCPCA 180
 QY 181 VDTAVSSSLVAVHLACGLRSGECSTALAGVSLMSTLTVMLSTKTPALARDCKAFS 240
 DB 181 VDTAVSSSLVAVHLACGLRSGECSTALAGVSLMSTLTVMLSTKTPALARDCKAFS 240
 QY 241 AEAQFGRGECQAVVYLKRLSGARADGRILAAVIRGSAIINDGASGLTVNGSSQELVL 300
 DB 241 AEAQFGRGECQAVVYLKRLSGARADGRILAAVIRGSAIINDGASGLTVNGSSQELVL 300
 QY 301 KRALADACAAASVGYVEAHGTGTLGPBIEIOLANAYVGLGRDVAETLLIGSVKTNLGH 360
 DB 301 KRALADACAAASVGYVEAHGTGTLGPBIEIOLANAYVGLGRDVAETLLIGSVKTNLGH 360
 QY 361 PEYASGITGLKLVLSLOHGOI PAHLAQAALNPKISKQDLRLTYTTRATTPMDWTPPRA 420
 DB 361 PEYASGITGLKLVLSLOHGOI PAHLAQAALNPKISKQDLRLTYTTRATTPMDWTPPRA 420
 QY 421 GVSSFGMSGTNAHVLEBAPAACTCPPAPEBPALVLSARTASALDQAARLDHLETY 480
 DB 421 GVSSFGMSGTNAHVLEBAPAACTCPPAPEBPALVLSARTASALDQAARLDHLETY 480
 QY 481 PSQCLGDVAFSLATTRSAMEHRLAVAATSRGLRAALDAAAGOTSFGAVSIDSRSRG 540
 DB 481 PSQCLGDVAFSLATTRSAMEHRLAVAATSRGLRAALDAAAGOTSFGAVSIDSRSRG 540
 QY 541 LAFLFTGGCAOTLGMKRGLYVWMSAFREAPDLCYRLPQOELDRPREVMAEPPASVDAL 600
 DB 541 LAFLFTGGCAOTLGMKRGLYVWMSAFREAPDLCYRLPQOELDRPREVMAEPPASVDAL 600
 QY 601 LDQTAFTOPALFTEFEYALALMRSMGVPELVAGHSIGELVAAQVAFSLEDVFLVAA 660
 DB 601 LDQTAFTOPALFTEFEYALALMRSMGVPELVAGHSIGELVAAQVAFSLEDVFLVAA 660
 QY 661 RGRMLQALPAGANVSIAPADVAAAAPPAASVIAAANAPQVYIAGAGOPVHAIAA 720
 DB 661 RGRMLQALPAGANVSIAPADVAAAAPPAASVIAAANAPQVYIAGAGOPVHAIAA 720
 QY 721 AMAARGARTKALVSHAFSPMLAMPLEAFERVAESYRRPSYLVNSLGGKACTDEVS 780
 DB 721 AMAARGARTKALVSHAFSPMLAMPLEAFERVAESYRRPSYLVNSLGGKACTDEVS 780
 QY 781 SPGVVVRHAREVVRPADGVKALHAAAGTFVEVBPKSTLGLVPCAMPDARPALIASRA 840
 DB 781 SPGVVVRHAREVVRPADGVKALHAAAGTFVEVBPKSTLGLVPCAMPDARPALIASRA 840
 QY 841 GRDEPATVLEALGGLMAVGLVSWAGLPPSGGRVPLPTYPMOBERYVIDTKADDAAGD 900
 DB 841 GRDEPATVLEALGGLMAVGLVSWAGLPPSGGRVPLPTYPMOBERYVIDTKADDAAGD 900
 QY 901 RRAFGAGHDEVEEGANVGGRRSARLDHPPESGRKRVNAAGDRPRLTIDEPGVLDH 960
 DB 901 RRAFGAGHDEVEEGANVGGRRSARLDHPPESGRKRVNAAGDRPRLTIDEPGVLDH 960
 QY 961 LVLRTERARRPGLGEVEIAVDAAGISFNDVOLALGMVDDLPKKNPPLLGGECAGRIY 1020
 DB 961 LVLRTERARRPGLGEVEIAVDAAGISFNDVOLALGMVDDLPKKNPPLLGGECAGRIY 1020
 QY 1021 AVGEVNGLVVGGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080
 DB 1021 AVGEVNGLVVGGPPIALISAGAFATHTTSAALVLRPQALSAIEAAMPVAYLTAMAYAL 1080
 QY 1081 DRIARLQGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAVLESIGVYSDS 1140
 DB 1081 DRIARLQGERVLIHAATGVGLAAVQAAQHVGAHVHATAGTPEKRAVLESIGVYSDS 1140
 QY 1141 RSDRVAADVRAVTGEGVDVNLNSGELIDKSFNLSHGRFVELGRDCCADNQLGR 1200
 DB 1141 RSDRVAADVRAVTGEGVDVNLNSGELIDKSFNLSHGRFVELGRDCCADNQLGR 1200
 QY 1201 PFLRLSFLVDLRGMLEPRPARVRLLELLGLIAAGVFTPPPIATLPPIARVADAFRSM 1260
 DB 1201 PFLRLSFLVDLRGMLEPRPARVRLLELLGLIAAGVFTPPPIATLPPIARVADAFRSM 1260

QY 1261 AQAQHLGKLVLTIGDPEVOQIRIPTHAGAGPSTGDRDLDRLASAAPARAALFAFLRTQ 1320
DB 1261 AQAQHLGKLVLTIGDPEVOQIRIPTHAGAGPSTGDRDLDRLASAAPARAALFAFLRTQ 1320
QY 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380
DB 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380
QY 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421
DB 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421

RESULT 5
US-09-568-486-2
; Sequence 2, Application US/09568486
; Patent No. 6358459
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1421
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-568-486-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRPIERAEDPIAIVGASGRLPGVYIDLSGFMTLLGSRDVTGRVPAERMDAAAFDP 60
DB 1 VADRPIERAEDPIAIVGASGRLPGVYIDLSGFMTLLGSRDVTGRVPAERMDAAAFDP 60
QY 61 DDDAPEGKTPVTRASFSLSDVACFDASFGISPREALRMDPAHRLILEVCWEALENAALAPS 120
DB 61 DDDAPEGKTPVTRASFSLSDVACFDASFGISPREALRMDPAHRLILEVCWEALENAALAPS 120
QY 121 ALVGTGTVFRTIGSPSEYEALPQATSAETDAHGIGTWPVSAGRISTALGRGCVA 180
DB 121 ALVGTGTVFRTIGSPSEYEALPQATSAETDAHGIGTWPVSAGRISTALGRGCVA 180
QY 181 VDTAASSSLVAVHLACOSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARDRCAPFS 240
DB 181 VDTAASSSLVAVHLACOSLRSGECSTALAGVSLMLSPSTLWMLSKTRALARDRCAPFS 240
QY 241 AADGFRGEGCAVVLKRLSGARADDRILAVIRGSAINHGDASSGLTVPNSSQEIIVL 300
DB 241 AADGFRGEGCAVVLKRLSGARADDRILAVIRGSAINHGDASSGLTVPNSSQEIIVL 300
QY 301 KRALADAGCAASSGVYEAHGTGTTLDPIEIOALNVYGLGRVATPPLIGVKTMLGH 360
DB 301 KRALADAGCAASSGVYEAHGTGTTLDPIEIOALNVYGLGRVATPPLIGVKTMLGH 360
QY 361 PEVASGITGLKLVLSLOHGOIPAHILAOALNPRISWGDRLITVTRARTPMDNTPRRA 420
DB 361 PEVASGITGLKLVLSLOHGOIPAHILAOALNPRISWGDRLITVTRARTPMDNTPRRA 420
QY 421 GVSSFGSGTNAHVLEAPPAATCTPPAPERPAELVLVSARTASALDQAARLDHLETY 480
DB 421 GVSSFGSGTNAHVLEAPPAATCTPPAPERPAELVLVSARTASALDQAARLDHLETY 480

DB 421 GVSSFGSGTNAHVLEAPPAATCTPPAPERPAELVLVSARTASALDQAARLDHLETY 480
QY 481 PSOCIGVAFSLATTSAMEHRLAUAATSEHGRRAALDAACQOTSFGAVRSIADSSRGK 540
DB 481 PSOCIGVAFSLATTSAMEHRLAUAATSEHGRRAALDAACQOTSFGAVRSIADSSRGK 540
QY 541 LAELFTGQAGTGLAMGRLVDVMSAPREAFDLCVRLPNOELDRPLREVMMAEPASVDAL 600
DB 541 LAELFTGQAGTGLAMGRLVDVMSAPREAFDLCVRLPNOELDRPLREVMMAEPASVDAL 600
QY 601 LDQTAFTQPALFTFEYALALMSWGPBVLVAGHSIGELVAAVCVAGVFSLEDAVFLVAA 660
DB 601 LDQTAFTQPALFTFEYALALMSWGPBVLVAGHSIGELVAAVCVAGVFSLEDAVFLVAA 660
QY 661 RGRIMQALPAGGAMVSIEAPEADVAAVAPHAASVSTIAAVNAPDOVVIAGAGOPVHATA 720
DB 661 RGRIMQALPAGGAMVSIEAPEADVAAVAPHAASVSTIAAVNAPDOVVIAGAGOPVHATA 720
QY 721 AMARAGARTKALHVSHAFFSPLMAPLEAFGRVAESVSYRPSIVLSNLSGKACTDEVS 780
DB 721 AMARAGARTKALHVSHAFFSPLMAPLEAFGRVAESVSYRPSIVLSNLSGKACTDEVS 780
QY 781 SPGYWRHAAEVYRFADGVKALHAAGGTVEVGPSTLLGLVPACMPDARPALASSRA 840
DB 781 SPGYWRHAAEVYRFADGVKALHAAGGTVEVGPSTLLGLVPACMPDARPALASSRA 840
QY 841 GRDEPATVLEALGSLMAVGLVSMAGLFPSSGGRVPLPTYPMOBERYWDITKDDAARGD 900
DB 841 GRDEPATVLEALGSLMAVGLVSMAGLFPSSGGRVPLPTYPMOBERYWDITKDDAARGD 900
QY 901 RRAAGAGHDEVEEGAVRGDRRSARLDHPPESSGRREKVEAAGDRPFLIEDPGVLDH 960
DB 901 RRAAGAGHDEVEEGAVRGDRRSARLDHPPESSGRREKVEAAGDRPFLIEDPGVLDH 960
QY 961 LVNRTERRAPGIGVEYIANDAGLSFNDVQLAGVPPDLCKRPNPLLLGEGCAIRIV 1020
DB 961 LVNRTERRAPGIGVEYIANDAGLSFNDVQLAGVPPDLCKRPNPLLLGEGCAIRIV 1020
QY 1021 ANGEVNGLVVGOPTIALSAGAFATHTTSAALVLPPOALSIAEAAAFVAYLTAVYAL 1080
DB 1021 ANGEVNGLVVGOPTIALSAGAFATHTTSAALVLPPOALSIAEAAAFVAYLTAVYAL 1080
QY 1081 DRIARLOPGERVLIHAATGVGLAAVQMAQHVGAEVHATGTEPKARYLESIGRVYSDS 1140
DB 1081 DRIARLOPGERVLIHAATGVGLAAVQMAQHVGAEVHATGTEPKARYLESIGRVYSDS 1140
QY 1141 RSDRFVADVATMTGSGVDVYVLSLSELIDKSFNLRSHGRFVELEGKRCYADNOLGLR 1200
DB 1141 RSDRFVADVATMTGSGVDVYVLSLSELIDKSFNLRSHGRFVELEGKRCYADNOLGLR 1200
QY 1201 PFLRNLSFSLVDLRGMWLEBPVRLALBELGLIAAGVTPPIATLPTARVADARSM 1260
DB 1201 PFLRNLSFSLVDLRGMWLEBPVRLALBELGLIAAGVTPPIATLPTARVADARSM 1260
QY 1261 AQAQHLGKLVLTIGDPEVOQIRIPTHAGAGPSTGDRDLDRLASAAPARAALFAFLRTQ 1320
DB 1261 AQAQHLGKLVLTIGDPEVOQIRIPTHAGAGPSTGDRDLDRLASAAPARAALFAFLRTQ 1320
QY 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380
DB 1321 VSQVLRTPPEIKVGAELFTRLGMDSLMAVELRNRIEASLKLKSTTFLSTSPNIALAON 1380
QY 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421
DB 1381 LLDALATALSLEVAENLRAGVONDFVSSGADQDMEIIAL 1421

RESULT 6
US-09-568-472-2
; Sequence 2, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schnupp, Thomas

APPLICANT: Ligon, James
APPLICANT: Molnar, Jecvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,472
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 1421
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-568-472-2

Query Match 100.0%; Score 7210; DB 4; Length 1421;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRIEAAEDPIAIVGASCRIPGVITLGGFTLLESGRDTVGRVPAERMDAAAFDP 60
DB 1 VADRIEAAEDPIAIVGASCRIPGVITLGGFTLLESGRDTVGRVPAERMDAAAFDP 60
QY 61 DPDAKPTVTRASFVAVACPDASFPGISPREALRMDPAHRLLEVCWEALENNAIAPS 120
DB 61 DPDAKPTVTRASFVAVACPDASFPGISPREALRMDPAHRLLEVCWEALENNAIAPS 120
QY 121 ALVGEVGVIGIGISEYEALPOATSAEIDAHGGLTGPSVGARISYALGRGPCVA 180
DB 121 ALVGEVGVIGIGISEYEALPOATSAEIDAHGGLTGPSVGARISYALGRGPCVA 180
QY 181 VDTAVSSSLVAVHLACQSLRSCECTALAGVSLMSPSTLWMLSKTRALARDGRCKAFS 240
DB 181 VDTAVSSSLVAVHLACQSLRSCECTALAGVSLMSPSTLWMLSKTRALARDGRCKAFS 240
QY 241 AADGFGEGCAVAVVTKRLSGARADGRILAVIRGSAIINHDCASSGLTPNGSQEIVL 300
DB 241 AADGFGEGCAVAVVTKRLSGARADGRILAVIRGSAIINHDCASSGLTPNGSQEIVL 300
QY 301 KRALADAGCAASVGVYEAHGTGTPPIEIOANAYVGRDVAATPLLIGSVYTNLGH 360
DB 301 KRALADAGCAASVGVYEAHGTGTPPIEIOANAYVGRDVAATPLLIGSVYTNLGH 360
QY 361 PEYASGITGLKVVLSLQHGOI PAHIAAOALNPRISMGLRLTVTRARTPMDWTPPRA 420
DB 361 PEYASGITGLKVVLSLQHGOI PAHIAAOALNPRISMGLRLTVTRARTPMDWTPPRA 420
QY 421 GVSSFGMSGTAHVLEBAPAACTCPAPERPABELIVSARTASALDQAARLRDLHETTY 480
DB 421 GVSSFGMSGTAHVLEBAPAACTCPAPERPABELIVSARTASALDQAARLRDLHETTY 480
QY 481 PSOCIGDVAFLATTRSMERHLAAATSRBSGLRALDAAAGGSPAVASIASSSGK 540
DB 481 PSOCIGDVAFLATTRSMERHLAAATSRBSGLRALDAAAGGSPAVASIASSSGK 540
QY 541 LAFLFTGGAOTLIGWGRGLYDVWSAFREAFDLCVLFNOELDRPLREVMMAEPASVDAL 600
DB 541 LAFLFTGGAOTLIGWGRGLYDVWSAFREAFDLCVLFNOELDRPLREVMMAEPASVDAL 600
QY 601 LDQTAFTOPALFTFEYALAAALMRSGVPELVAAGHSISELVAAACVAFSLEDVAVLAA 660
DB 601 LDQTAFTOPALFTFEYALAAALMRSGVPELVAAGHSISELVAAACVAFSLEDVAVLAA 660
QY 661 RGRIMQALPAGAMWISAPREADVAAPPAASISIANVAPDOVVAGGQPHATIA 720
DB 661 RGRIMQALPAGAMWISAPREADVAAPPAASISIANVAPDOVVAGGQPHATIA 720
QY 721 AMAARGARTKALHVSFHSPLMAPMLEAFGRVAESVYRRPSIYLVNLSGKACTDEVS 780
DB 721 AMAARGARTKALHVSFHSPLMAPMLEAFGRVAESVYRRPSIYLVNLSGKACTDEVS 780

DB 721 AMAARGARTKALHVSFHSPLMAPMLEAFGRVAESVYRRPSIYLVNLSGKACTDEVS 780
QY 781 SPGYWVRHAREVRFPAGVYKALHAAGAGTFVEVCPKSTLLGLVAPCPDPARPALLASSRA 840
DB 781 SPGYWVRHAREVRFPAGVYKALHAAGAGTFVEVCPKSTLLGLVAPCPDPARPALLASSRA 840
QY 841 GRDEPATVLEALGMLAVGVLSWAGLFPSCGRVPPPTYPMQBERWIDTKADDAARGD 900
DB 841 GRDEPATVLEALGMLAVGVLSWAGLFPSCGRVPPPTYPMQBERWIDTKADDAARGD 900
QY 901 RRABAGHDEVEEGCAVAGGDRSARLDHPPESSGREKVEAAGDRPRLIDEVGLDH 960
DB 901 RRABAGHDEVEEGCAVAGGDRSARLDHPPESSGREKVEAAGDRPRLIDEVGLDH 960
QY 961 LVAFVTERRAAGLGEVEIYDAAGLSFNDVQALGMPDDLPGRPNPPLLGGSCACRIV 1020
DB 961 LVAFVTERRAAGLGEVEIYDAAGLSFNDVQALGMPDDLPGRPNPPLLGGSCACRIV 1020
QY 1021 AVGEGVGLVVGQVPIALSACAFTHVTTSAALVLPPOALSAIEAAMPVAVLTAWYAL 1080
DB 1021 AVGEGVGLVVGQVPIALSACAFTHVTTSAALVLPPOALSAIEAAMPVAVLTAWYAL 1080
QY 1081 DRIARLOPGERVLIHAAATGGVGLAAVQAOHVGAENATAGTPEKRAVLESIGVRYVSDS 1140
DB 1081 DRIARLOPGERVLIHAAATGGVGLAAVQAOHVGAENATAGTPEKRAVLESIGVRYVSDS 1140
QY 1141 RSDRFVADVARTGEGVDVYVLSLSEGLIDKSNLRSRFRFELGKRDCAVANOGLR 1200
DB 1141 RSDRFVADVARTGEGVDVYVLSLSEGLIDKSNLRSRFRFELGKRDCAVANOGLR 1200
QY 1201 PFLNLSFSLVDLGMMLERPARVALLLEGLIAAGVFTPPPIATLPIARVADAFRSM 1260
DB 1201 PFLNLSFSLVDLGMMLERPARVALLLEGLIAAGVFTPPPIATLPIARVADAFRSM 1260
QY 1261 AQAHLGKLVITLGDPEVOIRIPTHAGAGPSTGDRLLDRLASAPAPARAALAEFLRTQ 1320
DB 1261 AQAHLGKLVITLGDPEVOIRIPTHAGAGPSTGDRLLDRLASAPAPARAALAEFLRTQ 1320
QY 1321 VSOVLRPEIRVGAELFTRIGMDSIMAVEIRNIEASLKKSTTFLSTSPNIALAQN 1380
DB 1321 VSOVLRPEIRVGAELFTRIGMDSIMAVEIRNIEASLKKSTTFLSTSPNIALAQN 1380
QY 1381 LLDALATALSERVAENLRAGVONDFVSSGADQDWEIHAL 1421
DB 1381 LLDALATALSERVAENLRAGVONDFVSSGADQDWEIHAL 1421

RESULT 7
US-09-567-899-2
Sequence 2, Application US/09567899
Patent No. 6383787
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Jecvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,899
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 1421
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-567-899-2
Query Match 100.0%; Score 7210; DB 4; Length 1421;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 VADRIEPAADPIAIVGASCRLLPGGVIDLSGFWTLLEGSRTVGRVPAERMDAAAMPDP 60
Db 1 VADRIEPAADPIAIVGASCRLLPGGVIDLSGFWTLLEGSRTVGRVPAERMDAAAMPDP 60
QY 61 DPDAAGKTPVTRASFSLDVACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIAPS 120
Db 61 DPDAAGKTPVTRASFSLDVACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIAPS 120
QY 121 ALVGTETGVFTIGIGSEYEALPQATASAEIDAHGCLTMSVSGARISYALGIRPCVA 180
Db 121 ALVGTETGVFTIGIGSEYEALPQATASAEIDAHGCLTMSVSGARISYALGIRPCVA 180
QY 181 VDTAYSSSLVAVHLACOSLRSGECTALAGVSLMLSTSTVWLSKTPALARDCKAFS 240
Db 181 VDTAYSSSLVAVHLACOSLRSGECTALAGVSLMLSTSTVWLSKTPALARDCKAFS 240
QY 241 AEADGFGREGCAVVLKRLSGARADGDRILAVIRGSAIINHGAASSGLTVENGSSQEIYL 300
Db 241 AEADGFGREGCAVVLKRLSGARADGDRILAVIRGSAIINHGAASSGLTVENGSSQEIYL 300
QY 301 KRALADAGCAASVGYVEAHGTGTTLDPIEIOALNAVYGLGRDVAITPLLIGSVKTNLGH 360
Db 301 KRALADAGCAASVGYVEAHGTGTTLDPIEIOALNAVYGLGRDVAITPLLIGSVKTNLGH 360
QY 361 PEYASGITGLKVLSTIOHGOIPALHQAOLNPRISSMDLTLTVRATPMPDMNTPRRA 420
Db 361 PEYASGITGLKVLSTIOHGOIPALHQAOLNPRISSMDLTLTVRATPMPDMNTPRRA 420
QY 421 GVSSFGMSGTAAHVLEBPAPATCTPPAPERPAELLVLSARTASALDQAARLRDHLTTY 480
Db 421 GVSSFGMSGTAAHVLEBPAPATCTPPAPERPAELLVLSARTASALDQAARLRDHLTTY 480
QY 481 PSQCLGDVAFLATTRSAMEHRLAVALTSREGLRALDAAAOQGTSPGAVSISADSSRGX 540
Db 481 PSQCLGDVAFLATTRSAMEHRLAVALTSREGLRALDAAAOQGTSPGAVSISADSSRGX 540
QY 541 LAFIFGTGGAGTTLGMRGLYDVWSAFREAPFLCYRLFNOELDRPRETMMAPASVDAL 600
Db 541 LAFIFGTGGAGTTLGMRGLYDVWSAFREAPFLCYRLFNOELDRPRETMMAPASVDAL 600
QY 601 LDQTAFTQPALFTFEYALALMRSGVPELVAGHSISELVAACTYAGVFSLEDAVFLVA 660
Db 601 LDQTAFTQPALFTFEYALALMRSGVPELVAGHSISELVAACTYAGVFSLEDAVFLVA 660
QY 661 RGRIMQALPAGGAMVSIAPADVAVAAPHAASVSIANAAPDQVITAGAGQPVHAIATA 720
Db 661 RGRIMQALPAGGAMVSIAPADVAVAAPHAASVSIANAAPDQVITAGAGQPVHAIATA 720
QY 721 AMARAGARTKLAHSHAFSPMLAPMLEAFGRVAESVYRRPSITLVNLSGKACTDBVS 780
Db 721 AMARAGARTKLAHSHAFSPMLAPMLEAFGRVAESVYRRPSITLVNLSGKACTDBVS 780
QY 781 SPGYVWRHAREVFAFDGVKALHAAGACTFVEVCPKSTLLGLVPACMDAPALLASSRA 840
Db 781 SPGYVWRHAREVFAFDGVKALHAAGACTFVEVCPKSTLLGLVPACMDAPALLASSRA 840
QY 841 GRDEPATVLEALGGLMAVGLVSNAGLPSSGGRVPLPTYPWQREYWIOTKADDAAGD 900
Db 841 GRDEPATVLEALGGLMAVGLVSNAGLPSSGGRVPLPTYPWQREYWIOTKADDAAGD 900
QY 901 PRAAGAGDEVEEGGAVRGDRRKARLHPPESSGRKKEVNAADRPRLSIDERGUTDH 960
Db 901 PRAAGAGDEVEEGGAVRGDRRKARLHPPESSGRKKEVNAADRPRLSIDERGUTDH 960
QY 961 LVLRTERBARAGLGEVEIAVDAAGLSFNDVOLALGMVDDLPGRKPNPPLLGGECAGRIY 1020
Db 961 LVLRTERBARAGLGEVEIAVDAAGLSFNDVOLALGMVDDLPGRKPNPPLLGGECAGRIY 1020
QY 1021 AVGSGVGLVVGQPVIALSAGAFATHTVTSALVLPDQALSAIEAAMPVAYLTAWYAL 1080
Db 1021 AVGSGVGLVVGQPVIALSAGAFATHTVTSALVLPDQALSAIEAAMPVAYLTAWYAL 1080
```

```
Db 1021 AVGSGVGLVVGQPVIALSAGAFATHTVTSALVLPDQALSAIEAAMPVAYLTAWYAL 1080
QY 1081 DRIARLOGERVULHAATGGVGLAAVOMACHVGAEVHATAGTPEKRAVLESIGRYVSDS 1140
Db 1081 DRIARLOGERVULHAATGGVGLAAVOMACHVGAEVHATAGTPEKRAVLESIGRYVSDS 1140
QY 1141 RSDRFVADVRAWTGEGVDVVLNSLSGELIDKSFNLLRSRSHRPYELGKRDICYADNOGLR 1200
Db 1141 RSDRFVADVRAWTGEGVDVVLNSLSGELIDKSFNLLRSRSHRPYELGKRDICYADNOGLR 1200
QY 1201 PFLRNLSFSLVDLGMMLERPARVALLBEELGLIAAGVFPPIATPIPARVADAFRSM 1260
Db 1201 PFLRNLSFSLVDLGMMLERPARVALLBEELGLIAAGVFPPIATPIPARVADAFRSM 1260
QY 1261 AQAHLGLVLTLDPEVQIRIPTHAGAGPSTGRDLDRLASAAPARAALAEFLRTQ 1320
Db 1261 AQAHLGLVLTLDPEVQIRIPTHAGAGPSTGRDLDRLASAAPARAALAEFLRTQ 1320
QY 1321 VSOVLRTPEIKVGAELFTRLGMDSLMAVELRNRIEASLKUKLSTTFLSTSPNIALAQN 1380
Db 1321 VSOVLRTPEIKVGAELFTRLGMDSLMAVELRNRIEASLKUKLSTTFLSTSPNIALAQN 1380
QY 1381 LLDLALATLSLIERVAENLRAGVNDVYSGADDDWEIHAL 1421
Db 1381 LLDLALATLSLIERVAENLRAGVNDVYSGADDDWEIHAL 1421
```

```
RESULT 8
US-09-335-409-7
; Sequence 7, Application US/09335409.
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Jeltvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPROTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
; US-09-335-409-7

Query Match 57.4%; Score 4137; DB 3; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8.5e-287;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;
```

```
QY 6 IERAADPIAIVGASCRLLPGGVIDLSGFWTLLEGSRTVGRVPAERMDAAAMPDPDPAP 65
Db 26 LEQERTPIAIVGICRFPFGADTPPEAFWELDSGRDAV--QPIDRKALVGNPSEVP 83
QY 66 GKTPTVTRASFSLD-VACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIAPSALVG 124
Db 84 -----RWAAGLITBAVDFDAAFGTSPREARSIDPQORLLLEVWEGLEDAGIAPQSIDG 138
QY 125 TENGVPFICIGISEYEALPQATASAEIDAHGCLTMSVSGARISYALGIRPCVAYDTA 184
Db 139 SRTGVFLGAGSDSYSHTVAAQORE-EGDAVDITENTLSVAAGRISTYTLGIGPCLTVDTA 197
QY 185 YSSSLVAVHLACOSLRSGECTALAGVSLMLSTSTVWLSKTPALARDCKAFSAEAD 244
Db 198 CSSSLVAVHLACRSLRARESDIALAGVNMLSKTMIMGLRIQALSPDGCRFRFDASAN 257
QY 245 GFGREGCAVVLKRLSGARADGDRILAVIRGSAIINHGAASSGLTVENGSSQEIYLKRAL 304
Db 258 GFVRGEGCGMVLKRLSDAQRHGDRIWALIRGSAMNODGSRSTGIMANVLVLAQALREAL 317
```

QY 305 ADAGCAASVGVYEAHGTCTTIGDPIEIOALNAVYGLGDVATPILLIGSVKTNLGHPEYA 364
Db 318 OSARVDAGIIGVETHTGTSTLGDPIEVEALRAVIGPADAGSRCTGAVKTNLHLBDA 377
QY 365 SGITGLKVLVSLGQGPALHAQAALNPRISMGLRLVTTRAPRPPMPMTPRAGVSS 424
Db 378 AGVAGLIRKALALHHELLPRNHPHTLNPRIRIESTALALATEPRPMPAGPRRAGVSA 437
QY 425 FGMSTNAHVLEAPPAATCTPPAPERPAELLVLSARTASALDAQAARLDHLETTYPQC 484
Db 438 FLSGTNVHVLEAPATVATPATPCRSALLVLSKSAALDAQAARLSAHIAAYBEQG 497
QY 445 LGDVAFSLATTSAMEHRLAVALAATREGIRALDLDAAGOTSPGAVRSIADSSRKLAVL 544
Db 498 LGDVAFSLATTSAMEHRLAVALAATREGIRALDLDAAGOTSPGAVRSIADSSRKLAVL 557
QY 545 FTGQAGTIGMGRGLYDVMSAFREAFDLGVRLFNQELDRPLREVMMAEPASVDALLDOT 604
Db 558 FAGQAGVPGMGRGLMEAMPAREFTFDRCVTLFDEELHOPLCBVMMABEGSSRLDQT 617
QY 605 AFTQAPALFTEVYALALMWSGVEBELVAGHSIGELVAACVAGVSLBDVFLVARGEL 664
Db 618 AFTQAPALFTEVYALALMWSGVEBELVAGHSIGELVAACVAGVSLBDVFLVARGEL 677
QY 665 MOALPAGGAMVSIAPADVAALVAAPHAASVIAAVNAPDQVYIAGQCPVIAIAAAMA 724
Db 678 MOALPAGGAMVSIAPADVAALVAAPHAASVIAAVNAPDQVYIAGQCPVIAIAAAMA 737
QY 725 RGARTKALVSHAFSPPLMAFMLFAFGVAESVSRPSIVLVSNLGKACTDEVSPPGY 784
Db 738 RGARTKALVSHAFSPPLMAFMLFAFGVAESVSRPSIVLVSNLGKACTDEVSPPGY 797
QY 765 WVRHAREVVRFAADVAYALHAAGAGTFVEYGPSTLGLVAPACPDARPALASSPAGDE 844
Db 798 WVRHAREVVRFAADVAYALHAAGAGTFVEYGPSTLGLVAPACPDARPALASSPAGDE 857
QY 845 PATVLEALGELMAVGLVSMAGLFPSCGRRVPLPYPMQREKRYM----DTKADDAAR-- 898
Db 858 AASALEALGELMAVGLVSMAGLFPSCGRRVPLPYPMQREKRYM----DTKADDAAR-- 917
QY 899 ----- 898
Db 918 AGGHPLGEVFSVTHAGLRMLWETLLDRKRLPMGBEHRQGEVFRGAGYLEMALSSGAE 977
QY 899 -----GDRAP----- 904
Db 978 ILGDGPQVTVVLLIETLTFAGDTAVPVQVVTTEERPGRLPQVASREGERRAPFRIHA 1037
QY 905 ----- 904
Db 1038 RGVLRIRIGVETTPASNTLALFARLHAAPVAPAAIYGALAEWGLQYGPALRGIAELMRGEG 1097
QY 905 ----- 904
Db 1098 BALGRVRLPEAAGSATAYQLHPVLLDACVQMTVGAFAORDDEATPMAPVEVGSVRLFORSP 1157
QY 905 ----- 904
Db 1158 GELMCHAVSDGOASSSRMSADPELMDGTGAVVAEISRLVVERLASGVRRRADDWPLE 1217
QY 905 -----GAGH----- 908
Db 1218 LDMEPALGPKITTAGRWLLGEGGGLGRSLCSALKAGHVVVHAAGDOSTTAGKRALIA 1277
QY 909 ----- 908
Db 1278 NAFDQAPTAVVVHSSLDGGGGLGPGLAGALDAPRSPDVADALLESALMRGCDVSL 1337
QY 909 ----- 908
Db 1338 VOALVGMDLRNAPLMLLTRGAQAAAAAGDVSVVQAPLLGLGRTIALHAEALRCISVDLP 1397

QY 909 -----DEVEGAVRGDRRSARLDHPPPSGREKVEAAGDRPFRL 952
Db 1398 AEPGEADALIAELLADDAEERVALRGDRVLAALVHRLPAQOREKEVEPPGDRPFRL 1457
QY 953 DEPGVLDLVLVTERARAPGLGEVEIAVDAAGLSFNVDQALGAVPDDLPGKPNPPLL 1012
Db 1458 DEPGVLDLVLVTERARAPGLGEVEIAVDAAGLSFNVDQALGAVPDDLPGKPNPPLL 1517
QY 1013 GECAIRYAVBEGVNGVLVGPVIALSAGAFATHTTSALVLPPOALSIAEAAAMVA 1072
Db 1518 SECAGRIYAVEGVNGVLVGPVIALSAGAFATHTTSALVLPPOALSIAEAAAMVA 1577
QY 1073 YLTAMVALDRIAPLOPGERVLIHAATGGVGAQVMAQVGAHVATAGTEPKAYLESL 1132
Db 1578 YLTAMVALDRIAPLOPGERVLIHAATGGVGAQVMAQVGAHVATAGTEPKAYLESL 1637
QY 1133 GVRVYSDRSRPFVADVAAMTGCGEVDVNLNSGELIDKSFNLRSHGRFVLEKRDY 1192
Db 1638 GVRVYSDRSRPFVADVAAMTGCGEVDVNLNSGELIDKSFNLRSHGRFVLEKRDY 1697
QY 1193 ADNOLGRLPPLRNTSFSGLVDLRGMLEPAPVRLLEELGLIAAGV----- 1239
Db 1698 ADNOLGRLPPLRNTSFSGLVDLRGMLEPAPVRLLEELGLIAAGV----- 1757
QY 1240 FTTPPIATLPIARVADAFRSMAOQHGLVTLTGDPPEVOIRIPHA----- 1286
Db 1758 FTTPPIATLPIARVADAFRSMAOQHGLVTLTGDPPEVOIRIPHA----- 1817
QY 1287 -----GAG----- 1289
Db 1818 GGLGLGLRVAGWMLAERGAGGLVLRSGAASABORAAVALAENHARVAVAKADVADS 1877
QY 1290 ----- 1289
Db 1878 QIERVAREVTASGMPRLRGVHAAGLVBDGLMOQTPAFRFTVMGPKVQGLHLHTLREA 1937
QY 1290 ----- 1289
Db 1938 PLSFVVLVYASAGLFGSGQGNVAAANAFDLALSHHRAQGLPALSDMGFTVGMAVA 1997
QY 1290 -----BST----- 1292
Db 1998 QENRGARQISRGKRGITPDEGLSALARLLEGDRVQTVIPIPPROWEFYATAASRLS 2057
QY 1293 -----GDRPLDRLASAPAAALAEFLRQVQVARTPEIKYGAELF 1338
Db 2058 RLVTQRAVDRTAGDRDLQLASAEPSARAQLQDVVRQVSHVRLPEDIKEVDAPL 2117
QY 1339 TRLGMSLMAVELNRRIRIASLKLKSTFSTSPNIALAONL-DALATAL 1389
Db 2118 SSMGMSLSIELNRRIRIALGVAAPALGWTYPTVAAITRWLDDALVRL 2169

RESULT 9
US-09-568-102-7
; Sequence 7, Application us/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Jettan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568.102
; PRIOR APPLICATION NUMBER: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7


```

; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-102-7

```

Query Match	Score	DB 4	Length
57.4%	4137		2439

QY	1	IEBAADPLAIYAGSARLPGGYIDLSGPTLLLEGSROTVGVVPERHDDAAMHPDDPAD	65
Db	26	LEOERTEPILAIYIGIRPFGGADTDEPAEWELLDSDRAN--OPIDRRMALGVHPSEEV	83
QY	66	GKTPVTRASFISD-VACPDASFEGGISPREALRMDPAHLLLEVCMEALENNALAPSLV	124
Db	84	-----RMAULLTEAVDGFDAAPFGISPREASLPDQCLLLEVWEGLDAGIAPGSLD	138
QY	125	TETGVFIGIPSEYEALPQATASAEIDAHGGLGTPMSVGARISYALGRGPCVAVDY	184
Db	139	SRFGVFLGACSSDYSHTVTAQORRE-EODAYDITGNTLSSVAGRLSYTLGLQGPCLTVDTA	197
QY	185	YSSSLVAHVLAQOSLRSGECSTALAGVSLMSPESTLWMLSKTALARDRCRFSLEAD	244
Db	198	CSSSLVAIHLACRSIRARESDIALAGVNMLLSSKXTMTIGRIQALS PDHCRFPDASAN	257
QY	245	GFGRGEGCAVVYLKRLSGARADGRIILVIRGSAIINHOGASSGLTVPNGSSOEIVLRAL	304
Db	258	GFPRGEGCGMVTYKRLSDAQHRGDIYMLIKGSAMNQDGRSTGLMAENVYLAQELRLREAL	317
QY	305	ADAGCAASSVGVEAHGTTTIGDPIETQALNAVYGLGRDVAETPLLIGSVKTMGHEVA	364
Db	318	QSRARVDAGAIQVETHTGTSIGDPIEVEALRAVLGPBARADSGRCVGAVTNMGHEGA	377
QY	365	SGTIGLLKTVLSLQHQCPAHLHAQALNPRISMGEDLRTYTRATTPMDMTPRAGVSS	424
Db	378	AGVAGTILKALMLHHEILLPRNLHFTLLPRIRIETGATLALTEBEPWMPRAGRPEFAGVSA	437
QY	425	FGMSGTNAVHVLEBAPATCTPAPBEPBELLVTSARTASALDAQAARLBDHLETPSQ	484
Db	438	FGLSGTNAVHVLEBAPATVLA PATGRSANELLVLSAKSAALDQAARLSGHTAAAYEG	497
QY	485	LGDVAFSLATTSAMEHRLVAATSBEGRLAALDAQAQOTS PGAVSS IADSSRGKIAFL	544
Db	498	LGDVAFSLVSTRSPMEHRLVAATSRREALRSLAEVAQOTPAGAAAGRAASSPGKIAFL	557
QY	545	FTGGGQOTIGMGRGLYDVWSAFREAFDLCYVLFENGLDRPLREYMWAEPAVDAAILLD	604
Db	558	FAQGGQVQVGMGRGLMEAMPAREFDFDCYVLFPEHLQPLCEYMWAEPPSSRSLLD	617
QY	605	AFTQPALFTFEYKALAAALMRMGVEPELVAHGSIGELVAAVCAVGFSLJEDVAVFLAARG	664
Db	618	AFTQPALFTLEVALAALFRSKGVEPELVAHGSIGELVAAVCAVGFSLJEDAVRLVVARGL	677
QY	665	MQALPAGAMVSIAPBADVAAA VAPHAASVIAAVNAPDOVITAGAGOPVHIAAAMA	724
Db	678	MQALPAGAMVSIAPBADVAAA VAPHAALVSIAAVNPBEQVITAGAEKFFQOIAAFPA	737
QY	725	RGARTKALVSHAFHSPPLMAPMLTA FGRVAASVSTRPSYLVSNLSGKACTDDEVSPGY	784
Db	738	RGARTKPLVSHAFHSPPLMDPMLTA FRRVETESVYTRPSIALVSNLSGKCTDEVSPAGY	797
QY	785	WYVHAREVVRPADGVYALDAAGAFVEVGPSTLLGLVPAQMPDARPALAASRAGRDE	844
Db	798	WYVHAREAVRFPDGYKALHDAAGAFVEVGRKPTLLGLVPCLDAPAPVLLPMSRAGRDE	857
QY	845	PATVEALGGLMAVGGVLSMAGLFPSSGRRVPLPTYPWQERRYI-----DTKADDAAR--	898
Db	858	AASALEALGFWVVGGSVTWSGVFPSSGRRVPLPTYPWQERRYIAPVVDREADGTGRAR	917
QY	899	-----898	
Db	918	AGHPLLGGEVFSVSTHAGILRWETLLDRKRLPWIGEHRAQGEVVPAGAYILENALSGAE	977

[illegible]

Db 2058 RLVTTORAVADRTAGDRDLEQLASBPBARAGLQDVVRVOVSHVLRKRPEDKIEVDAL 2117
Qy 1339 TRLGNDLSMAVELRNRIEASLKLTSTFLSTSPNIALIAQNL-DALATAT 1389
Db 2118 SSMGMDISLMSLELRRIEALGVAAAPALGWTPTVAATIRWLIDALVRL 2169

RESULT 10
US-09-567-969-7
Sequence 7, Application US/09567969
Patent No. 6355457
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,969
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 7
LENGTH: 2439
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-567-969-7

Query Match 57.4%; Score 4137; DB 4; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8.5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

Qy 6 IERAEDPAIVAGSACRLPGGVIDISGFWTLEGSRDYGRVAPARWDAAPFDDPAP 65
Db 26 LEQERTEPALVIGICRPPGADTPBPAPFELLDSGDV--QPLDRMLVGVHPSSEVP 83
Qy 66 GKTPTRAFSLSD-VACFDASFFGISPREALNMPARLLLEVCWEALENAIAPSAV 124
Db 84 ----RMAGLTFEAVDGFPAFFGTSPPRARSLDPOQRLLLEVTEGLEDACIAQSDG 138
Qy 125 TETGVFIGPSEYEALPQATASARIDAHGGLGTMPVAGRISYALGRPCVAVDTA 184
Db 139 SRTGVFLGACSSDYSHVTAQQRRE-EQDAYDITGNTLSVAAGRLSTTGLQGPCLTVDTA 197
Qy 185 YSSSLVAVHLACQSLRSRGCTALAGVSLMSPSTIVLSTKRALAPGRKAKASLEAD 244
Db 198 CSSSLVAHLACRSIRARESDIALAGVNMILSSKTMILGRIOALSPDGHCRTPDASAN 257
Qy 245 GFGREGGAVVVKLSGKSGADDRILAVIRGSAINHDSAGLTVPNSSQOEVILKAL 304
Db 258 GFVREGGCVVVKLSLSDQQRHGRHIMALIRGSANQDGRSTGLMAPVTLAGEALLREL 317
Qy 305 ADAGCAASSVGEVYAGTGTTLGDPLEIQALNAVYGLGRDVA TPILIGSVKTNLCHPEYA 364
Db 318 QSARVDAGAGIGVETHTGTSTLGDPIEVALRAVGPARADSGRCVLGAVKTNLHLEGA 377
Qy 365 SGIITLVVLSLQGOI PAHLAQAALNPRISMGDLRLVTTPARPPMPWMTPRRAGVSS 424
Db 378 AGVAGLITKAALHHELLPRNLHFTLNPRIEGLTALATEPVWPAPAGPRRAGVSA 437
Qy 425 FGMSTNAHVVEEAPAACTCPAPAPERPAELLVLSTARTASALDAQAARLDLETPSQ 484
Db 438 FGLSGTNHVVEEAPATVLA PATPGRSALLVLAKSAALDAQAARLSAIIAAYPEBG 497
Qy 485 LGDVAFSLATTSAMEHRLAVALATSRBEGRALDAQAAGQTS PGAVRSTADSSRGKLAFL 544
Db 498 LGDVAFSLATTSAMEHRLAVALATSRREALRSLAEVAAAGQOTPAGAARGAASSPGKLAFL 557

Qy 545 FTGGAQOTLGMGRGLYDVMSAFREAFDLCVRLFNOELDRPLREVMMAEPASVADALLDOT 604
Db 558 FAGGAQVPGMGKRGIMEMPAFRETFDRCYTLFDBELHQLPCEVMMAEPGSSRLDLOT 617
Qy 605 AFTQPAFTFEYALAAALRWSGVEBELVAGHSIGELVAAVAGVFSLEDAVFLVARGRL 664
Db 618 AFTQPAFTFEYALAAALFRSWGVEBELVAGHSIGELVAAVAGVFSLEDAVFLVARGRL 677
Qy 665 MQALPAGAMVSIAPADVAANAAPHASVSIAPVAPDQVITAGAQOPHATIAAMA 724
Db 678 MQALPAGAMVSIAPADVAANAAPHALVSIAPVAPDQVITAGAQOPHATIAAMA 737
Qy 725 RGARTKALVSHAFHSPMAFMLEAFGRVAVSVSYRPSIYLVNSLSKACTDEVSPPGY 784
Db 738 RGARTKPLVSHAFHSPMLDPMLEAFRVYTSVYTRPSIALVNSLSKPCPTDEVSABGY 797
Qy 785 WVRHAREVVRPADGVKALHAAAGAGTFVEVGKSTLLGLVPACMPDARALLASSRAGDE 844
Db 798 WVRHAREVVRPADGVKALHAAAGAGTFVEVGKPTLLGLVPACLPDARVLLPASRAGDE 857
Qy 845 PATVLEALGLMANVGLVSNAGLPPSGRRVPLPTYPQORRYM----DTKADDAAR-- 898
Db 858 AASALEALGFMVVGVSVTWVGSPSGRRVPLPTYPQORRYMTEADVREADGTGRAR 917
Qy 899 ----- 898
Db 918 AGGHPILGEVSVSHAGLRIMETTLDRKRLPMLGEHRAQGEVVPFGAGYLEMALSSAE 977
Qy 899 -----GRRAP----- 904
Db 978 ILGDGPQIVTDVLIETLTFAGDTRAPVQVVTTEERPGRLFOVASREPGERRAPFRJHA 1037
Qy 905 ----- 904
Db 1038 RGVLRIRIGRVETPARSNLALRARIHAAVPAAIYGALAENGLOYGALRGILAEIMRGE 1097
Qy 905 ----- 904
Db 1098 EALGRVRLPEAAGSATYATVQLPVLIDACVQIVGAFADRDENTPAVVEVSVLRFQSP 1157
Qy 905 ----- 904
Db 1158 GELMCHARVVDGQOASRMSADFELMDGTGAVVAEISRLVVERLSGVRRRDADWFLE 1217
Qy 905 -----GAGH----- 908
Db 1218 LDMEPALGCPKITAGRWLLIGGGGLGRSLCSALKAGHVVAHAAGDSTAGKALLA 1277
Qy 909 ----- 908
Db 1278 NAFDQAFYAVVHLSLDDGGQLGPGLAGAQAALDAPRSPVDADALBESALMRGDSVSL 1337
Qy 909 ----- 908
Db 1338 VOALVGMDLRNPRLMLITRGAQAAAAGDVSVQAPLLGLGRTTALBEHAEIRCSVDDP 1397
Qy 909 -----DEVEEGAVRGDRDSALDHPPESSGREGKEVTEAAGDRPFRLEI 952
Db 1398 AEPEDGEALLAEILLDAEEVVALRGDDRIVARIVHMLPAQORREKVEPGDDEFRLEI 1457
Qy 953 DEPGVLDLVTYRTERAPGLGEVEIAVDAAGLSFNDVQLALGWPPDLPCKPMPPLLG 1012
Db 1458 DEPGVLDLVTYRTERAPGLGEVEI SVEAAGLDSIDQLALGVAPNLPPEEBIEPLVIG 1517
Qy 1013 GECAGRIVAVEGVNGLVGGPVIALSAGAPATHTVTSALVLPPOALSAIEAAMFVA 1072
Db 1518 SECAGRIVAVEGVNGLVGGPVIALAAGVATHTVTSATVLPRLPILSATEAAMFVA 1577
Qy 1073 YLTAMVADLRARLQOPGRVULIHAATGVGLAAVOMAGVAGVHAATGTGTEKAYLESL 1132
Db 1578 YLTAMVADLRARLQOPGRVULIHAAGGVGLCAVRMAQVAGVAYATDTEENRYLESL 1637
Qy 1133 GVRVYSDRSRPFVADVAMTGGEGVDVNLISGELIDKSFNLIRSHGRFVEIGKRDY 1192

```

Db      1638 GVRVVSRSRSGRFYTDVAAWTDGEGVDVLDLSGERIDKSLMVLKAGRLVKGRDDC 1697
Qy      1193 ADNGLRPLRLNRSFSLVDLRGMMLEBRPARVRLLEEELGLIAGV----- 1239
Db      1698 ADTPGLPLRLNRSFSGVDLRGMMLDQPARLRALDELFGLVAAGAISPLGSLRGGS 1757
Qy      1240 FTTPPIATLPARVADAFRMAQAHLGKLVLTLDGPEVOGIRIPTHA----- 1286
Db      1258 LTPPEVEFPISRAAEFRMAAQGHGKLVLTLDPEVRIRAPRESSVAVRADOTYLV 1817
Qy      1287 -----GAG----- 1289
Db      1818 GGLGLGRVAGWLAERGAQVLVGRSGAASQAQRAAVALAAGARVAKADVADRS 1877
Qy      1290 ----- 1289
Db      1878 QIERVREVTAAGMPLRGVHAAGLVDDGLMOOTPARFRTWMBKVOGALHHTLREA 1937
Qy      1290 ----- 1289
Db      1938 PLSEFVLVYASAGLEFGSPGQGNVAAANAFDLALSHHRAQGLPALSDMKMFTVGMVA 1997
Qy      1290 -----PST----- 1292
Db      1998 QENRGARQISRGMGITPDEGLSALARLLEGGDVQGVIPITPRQWVEFYATAASRLS 2057
Qy      1293 -----GDRDLRLASAPARAPAAALEARTQVSGVLTPEIKYCAELF 1338
Db      2058 RLVTTPRAVDRTAGDRDLLEQLASAEPSARAGLLQDVAVQVSHVRLPEDKLEVDAPL 2117
Qy      1339 TRLGMDSLMAVELNRRIEASLKLKLTFTLSTPNIALLAQNLL-DALATAL 1389
Db      2118 SSMGMDSLMSLELNRIEALGVAAPALGWTYPTVVAITRWLLDDALVRL 2169

```

```

RESULT 11
US-09-568-480-7
; Sequence 7, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zickler, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,480
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-480-7

```

```

Query Match      57.4%; Score 4137; DB 4; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8,5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

```

```

Qy      6 TERAEDDIAIVGASCRLLPGVITDLSGFWTLLSGRDTVGRVPAERMDAAAFDDPDAP 65
Db      26 LEOETEDIAIVGICRPPGADTPAEPWELLDSGRDAV--QPLRRMALVGVHSEEV 83
Qy      66 GKTPTTASFLSD-VACEDASFFGISPREALRMDPAHLLLEVCHEALEMAIASALVG 124
Db      84 -----RWAGLLTEAVDGFDAAFGTSPRARSLDPOQRLLLEVTWEGLEDAGIAPOSIDG 138

```

```

Qy      125 TETGVFIGIPSEYEALPQATASAEIDAHGGLTMTSPVGAGRI SYALGLRGPCVAVDTA 184
Db      139 SRTGVFLGACSDSVSHYTAQORRE-EQDAYDITGNTLSVAAGRI SYTLGLQPCCLTDVTA 197
Qy      185 YSSSLVAVHLACQSLRSGEGSTALAGVSIMLSTVLMSTKTRALARDRCXAFSAEAD 244
Db      198 CSSSLVAIHILACRSLRREEDDLAAGVNMILSKRTIMIGRIQALSPPDGCRTFDASAN 257
Qy      245 GFGEGCAVVVLRKLSGARADGRILAVIRGASLHNDGASSGLTPNGSQEIVLKRAL 304
Db      258 GFVARGCGMVVLRKLDAGRHDRILWALIRGSMNDGSGTGLMARVLAQELLRAL 317
Qy      305 ADAGCAASVGYEBAHGTTGLDPIEIQALNAVYGLGRDVAIPLIGSVYTNLGHPEYA 364
Db      318 QSAVDAIGAIVYETHGTGTSLGPIEVEALRAVLGPADAGSRGCVLGAVTNLGHEGA 377
Qy      365 SGITGLLKVVLSTLOHGOIPAHILAAQALNPRISMGDDLTLTVTRATPMPDMNTPRRAG 424
Db      378 AGVAGLILKAALALHHEILPRILHFTLNPRIRIEGTALALATEPEVMPRAGRPRFAG 437
Qy      425 FGMGTNAHVVLLEAPATCTPPAPERPAELVLSTARTASALDQAARLDRHLETPSQ 484
Db      438 FGLSGTNHVVLLEAPATCTPPAPERPAELVLSTARTASALDQAARLDRHLETPSQ 497
Qy      485 LGDVAFSLATTRSAMEHRLAAVATSRGCLRAALDAAAQGTSPGAVSISADSSRGKLAFL 544
Db      498 LGDVAFSLVSTRSPMEHRLAAVATSRGCLRAALDAAAQGTSPGAGRAASPGKLAFL 557
Qy      545 FTGGCAOTLGMGRLLYDWSAFREAFDLCYRLFNOELDPLREVMMAEPASVDAALLDOT 604
Db      558 FAGGAQVPGMGRLLMEAMPAREFTFDRCVTLFRELHQPLCEVMMAEPGSSRSLLDOT 617
Qy      605 AFTOPALFTFEYALAAALMRSWGVPELVAGHSIGELVAAQVAGFSLDPAFLVLAARGL 664
Db      618 AFTOPALFALEYALAAALFRSGVPELVAGHSIGELVAAQVAGFSLDPAFLVLAARGL 677
Qy      665 MQALPAGAMVSIAPADVAVAAPHAASVSIANAAPDVVIAGAGQPVHAIAPAAVA 724
Db      678 MQALPAGAMVSIAPADVAVAAPHAALVSIANAANGPEQVVIAGAEKFAQIAAPAA 737
Qy      725 RGARTKALYSHASHSLPMLPMLPMLPMLPMLPMLPMLPMLPMLPMLPMLPMLPML 784
Db      738 RGARTKPLVSHASHSLPMLPMLPMLPMLPMLPMLPMLPMLPMLPMLPMLPMLPML 797
Qy      785 WVRHAREVVRPADVKVLLHAAGAGTPEVEGKSTLLGLVPCMPDAPRALASSRAGDE 844
Db      798 WVRHAREVVRPADVKVLLHAAGAGTPEVEGKSTLLGLVPCMPDAPRALASSRAGDE 857
Qy      845 PATVLEALGILMAVGLVSMAGLFPSSGRVPLPTYPQORERYI---DTKADDAAR-- 898
Db      858 AASALEALGFMVVGSGVTSWGVFPSSGRVPLPTYPQORERYI---DTKADDAAR-- 917
Qy      899 ----- 898
Db      918 AGHPHLLGEVFSVTHAGLRIMETTLDRKRLPWLGEHRAQGEVFPAGYLEMALSSGAE 977
Qy      899 -----GDRAP----- 904
Db      978 ILDGPIQVTDVLIETLTFRAGDTAVPVQVTTTEERFGRRLFOVASREPGERRAPFR 1037
Qy      905 ----- 904
Db      1038 RGVLRIRIGRVETPARSNLAAALRARLHAAPAAATYALAEWGLQYPALRGLAEUWREG 1097
Qy      905 ----- 904
Db      1098 EALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADREATPMAVGVSVLFORSP 1157
Qy      905 ----- 904
Db      1158 GELMCHARVVSDDQOASSRMSADPELMDGTGAVVAEISRLVVERKLASGVRRRDDDW 1217
Qy      905 -----GAGH----- 908

```

```

Db      1218 LDMEPALGSPKITAGRWLLLEGGGLGRSLCSALKAGHVVVHAAGDSTTAGKRALIA 1277
Qy      909 -----
Db      1278 NAFDQAPTAVVHLSLDGGGQLGPQLGAQALDAPRPDPVDADALLESALMGCDSTLSL 1337
Qy      909 -----
Db      1338 VQALVGMDLRNAPRLMLTRGAQAAAAAGVSVVQAPLLGLGRTIALEHAEICISVDLDP 1397
Qy      909 -----
Db      1398 AEPBEADALLAELADDAEVEALRGGRVLAARLVHRPDQORREKPEPADDRPFRLLEI 1457
Qy      953 DEPGVLDHIVLITERARGLCEVEIAYDAAGLSFNDVOLALGMVPPDLPGKPNPPLLIG 1012
Db      1458 DEPGALDQVLIRATGRAPGPEVEISVEAAGLDSIDIDIALGVAPNDLPGSEIEPLVIG 1517
Qy      1013 GECARIVAVGEGVGVVGVGVIALSAGAFATHTTSALVLRPPQALSAIEAAMPVA 1072
Db      1518 SECARIVAVGEGVGVVGVGVIALAAGVFATHTTSATLVLRPLGSAIEAAMPVA 1577
Qy      1073 YLTAVYALDRILARLPGERVILHAATGAGVLAQVQAHVGAEVHATAGTPEKRAYLESL 1132
Db      1578 YLTAVYALDKVAHLDAQGERVILHAENGAGVGLCAVRMAQVGAEVATADTPENRAYLESL 1637
Qy      1133 GVRVYSDSRSDRFVADVRAWTGEGEVVVVNSLSEGLIDKSFNLLRSHGRFVELGKRDY 1192
Db      1638 GVRVYSDSRGFRVVDVHAMTDEGVDVLDLSGERIDKSLMVLRACGLVKLRDLC 1697
Qy      1133 ADNOLGPRFENLSFSLVDLRGMMLERPARVALLLELGLIAGV----- 1239
Db      1698 ADTQGLPRLNLFNFSPOVDLRGMMLDQPARLRALDELFGVLAAGALSPLSGLRVGS 1757
Qy      1240 FTTPPIATLPILARVADAFRMAQAQHLKLVLTIGDPEVQIRIPTHA----- 1286
Db      1758 LTPPEVETPFIIRAEAPFRMAQOGHLGKLVLTIDDPVIRIAPBESSVAVADGTLYVT 1817
Qy      1287 -----GAG----- 1289
Db      1818 GGLGGLGLRVAGMLAERGAQQLVLRSGSAASEQRAVALEAGARVYAKADVADRS 1877
Qy      1290 ----- 1289
Db      1878 QIERVLEVTASGMPLRGVHAAGLVDDGLMQOTPARFRVTMGPRVQALHLATLTREA 1937
Qy      1290 ----- 1289
Db      1938 PLSPFVLVASAAGLRGSPGQGVYAAANAFDLASHHRAQGLRALSIDMGFTVGMAVA 1997
Qy      1290 -----PST----- 1292
Db      1998 QENRGARQISRGMRGITPDEGLSALARLEGGDRVOTGVPIRTPROWVEFYRPAATARRLS 2057
Qy      1293 -----GDRLLDLRLASAPARAALAEFLRTQVSQVLRTPBELKVGAEALF 1338
Db      2058 RLVTTORAVADRTAGDRDLLEQLASBPARGGLQDVVRVQVSHVLRIRPEKIEVDAPL 2117
Qy      1339 TRIGDSMLAVALVLRNRIEASLKLSTTFLSTSPNIALLAQNL-DALATAL 1389
Db      2118 SSMGDSLSLMSLELRRIEALGVAPALGMVTPYVAALTRWLDLDDALVVR 2169

```

RESULT 12
 US-09-568-486-7
 ; Sequence 7, Application us/09568486
 ; Patient No. 6355459
 ; GENERAL INFORMATION:
 ; APPLICANT: Schupp, Thomas
 ; APPLICANT: Ligon, James
 ; APPLICANT: Molnar, Istvan
 ; APPLICANT: Zirkle, Ross

```

; APPLICANT: Cyr, Devon
; APPLICANT: Goetlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568, 486
; CURRENT FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PR
; ORGANISM: Sorangium cellulosum
US-09-568-486-7

Query Match      57.4%; Score 4137; DB 4; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8.5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

Qy      6  TERAAEDPIAVGASCRLPGLGVYIDLSGFWTLLBGSRTDVGVRVPRMDAAMPDPDPAP 65
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      26  LEOETBEDIAIVIGICRPPGADTPEAFWELLDSGRDAV--QPLDRWALGVHPSEVP 83
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      66  GKTPTVTRASFUSD-VACPDASFPGISPREALRMDPAHRLLEVCMEALENAIAPALVG 124
Db      84  -----RWAGLITAEVDFGDAAPFTGTSPEAKSLDPQOHLLEVTWEGLEDGIAPQSLDG 138
Qy      125  TETGVFIGIGSEYEALPQATASAEIDAHGLGTMPBVGAGISYALGLRGPVAVDTA 184
Db      139  SRTGVFIGACSDSHVTAQQRRE-EDQAYDITGNTLSVAAGRLSYTLGLGPGCLTVDTA 197
Qy      185  YSSSLVAVHLCQSLRSECCSTALAGVSLMLSTVLMSTKALARDGCKAFSAEAD 244
Db      198  CSSSLVAVHLCRSLRAESDLAAGVNMLSKTMIMGRIGIALSPDGCRFPDASAN 257
Qy      245  GFGECEGAVVVLRLSGARADGRITAVIRGSAIMNDGASGGLTVPRGSGOEIVLRAL 304
Db      268  GFVRGEGGMVVLRLSLDQNRHGRIMALINGANNDGRSTGLMAPVVLAOEALLREAL 317
Qy      305  ADAGCAASSVGVVEANGITGLDPIEIOALNAVYGLGRDVATPLLIGSVKTNLGHPEYA 364
Db      318  OSARVDAGALGVETHTGTSIGDPIEVEALRAVLGPRADGSCVLCGVATNIGHLEGA 377
Qy      365  SGITGLKVLVSLQHQQLPAHLHAQALNPRISWGDRLTVTRATTPWDMNTPPRAGVSS 424
Db      378  AGVAGLIRAAALHHELIERNLHPTTNLRIRIEGTALALATEPVMPRAGRPRPAGVSA 437
Qy      425  FGMSGTNAHVLEAPATCTPPAPERPAFLVLSARTASALDQAARLRDHLLETYPSC 484
Db      438  FGLSGTNAVHLEAPATVLAIPATGRSAELLVLSAKSAAALDDQAARLSAHIAAYPEQ 497
Qy      485  LGDVAFLATTRSAMEHRLAVALATSRBGLRALDAAAGQOTSPGAIVSIADSSRGKALFL 544
Db      498  LGDVAFSLVSTRSPMEHRLAVALATSRBGLRALDAAAGQOTPGAARARAASSGCKALFL 557
Qy      545  FTGGAQQLGMRGLYDWSAFREAFDLCVLLFMQELDRPLREVMWMEPAVDAALDQOT 604
Db      558  FAGGAQVPMGMRGLMEWPAFRFDFRCVLLFQREHLQPLCEVWMAEPGSSRSLLDQOT 617
Qy      605  AFTQPALFTFEXALAAALMRSGVPELVAAGHSIGELVACVAGVSLDEDAFLVLAARGL 664
Db      618  AFTQPALALEYALAAALFRSGVPELVAAGHSIGELVACVAGVSLDEDAFLVLAARGL 677
Qy      665  MQALPAGAMVSIAPADVAANAAPHAAVSIANAAPDQVITAGACQPVHAIAMAANA 724
Db      678  MQALPAGAMVSIAPADVAANAAPHAAVSIANAAPDQVITAGACQPVHAIAMAANA 737
Qy      725  RGARTKALVSHAFRSPMAAPKLAFCGRVAAVSYSYRRSYLVNLSLSCAKTDEVSAGY 784
Db      738  RGARTKPLVSHAFRSPMDPMLAEFRVTSYVYRRSIALVNSLSKPCPTDEVSAGY 797
Qy      785  WVRHAREVVRPADVKALHAAGAGTFVEVGPKSTLLGLVPAACMPDARPALAASRAGRDE 844

```

```

Db 798 WVRAREAVRPADGKALHAAAGLFFVGVKPTLLGLVPACTLPDARVLLPASRAGDE 857
Oy 845 PATYLEALGMAVAGLVSMAGLFPSSGGRVLPFYMQRRYMT---DTKADDAAR-- 898
Db 858 AASALIEALGFMWVVGGSVTMSGVFPSSGRVLPFYMQRRYMTLEAVDEKADGTGAR 917
Oy 899 ----- 898
Db 918 AGHPHLLGEVPSVTHAGLRMETTLDRKRLPMIGEHRAQGVVFPAGVLEMLSSGAE 977
Oy 899 -----GRRAP----- 904
Db 978 ILGDGPLOVTVLLETITFAGDTAVPVQVYTTERRPGRLEFOVASREBERRAPFRITHA 1037
Oy 905 ----- 904
Db 1038 RGVLRRIGRVETPARSNLALPARLHAAPPAALYGALAENGLOYGALRGLAEIMRGE 1097
Oy 905 ----- 904
Db 1098 EALGRVRLPEAAGSATAYQHLPLLDACVMIVGAFAADREATPMAVENVGSVLFQKSP 1137
Oy 905 ----- 904
Db 1158 GELWCHARVVSDDGOASSRMSADELMDGTGVAVAEISRLVVERLASGVRRRDADWFL 1217
Oy 905 -----GAGH----- 908
Db 1218 LDWEPALGPKITAGRWLLIEGGGGLRSLCSALKKAQHVVVHAAGDSTAGKRALLA 1277
Oy 909 ----- 908
Db 1278 NAFPGQAPTAVVHLSLDDGGGQLGPGLGAGALDAPRPDVADALLESALMRGCDSVLSL 1337
Oy 909 ----- 908
Db 1338 VQALVGMDLRNAPRLMLTRGAQAAAAGVSVQAPLLGLRTTALAEHRLCISVDLP 1397
Oy 909 -----DEVEGAVRGGRSARLDHPPESSGREKVEAGDPRFLEI 952
Db 1398 AEPGEADALIAELLDAEEVVALRGSDRLVARLVHSLPAQOREKVEPGRDPRFLEI 1457
Oy 953 DEPGVLDHLVLRVTERRAPGLGEVEIANDAGLSFNVDQALGWPDDLPCKPNPPLLG 1012
Db 1458 DEPGALDQVLRATGRRAPGGEVEISVEAAGDSIDQLALGVAANDLPGEIEPLVLS 1517
Oy 1013 GECAGRIVAVEGVNGLVGPVIALSAGAFATHTTSSAALVLPPOALSIAEAAAFVA 1072
Db 1518 SECAGRIVAVEGVNGLVGPVIALAGVAFATHTTSSATLVLPPLGLSATEAAMPPLA 1577
Oy 1073 YLTAMVYALDIARLQOPGRVLIHAATGVGLAAYOMAGHGAEVAHTAGTEKAYLES 1132
Db 1578 YLTAMVYALDKVAHLQAGERVLIHAEAGGVGCAVARMQAVAEVATADTEENAYLE 1637
Oy 1133 GVRVYSDRSRDPFVADVAMTAGEGVDVVLNLSGELDKSFNLLRSHGRFVLEKRDY 1192
Db 1638 GVRVYSDRSRGRFTVDHAMTDEGVDVVDLSLSERIDKSLMWLRACGRVLEKGRDDC 1697
Oy 1193 ADNOLGRPLRLNLSFSLVLDLRGMWLERPARVALLLEELGLIAAGV----- 1239
Db 1698 ADTQGLPRLRLNLSFSGVDLRGMWLDQPARIRALLDELFGVVAAGALSPLSGLRVGS 1757
Oy 1240 FTPEPIATLPIAVADAFRSMQAQHLGLVLTGLDEPVQORIPRHA----- 1286
Db 1758 LTPPEVETPISRAAEARRMAQGOHLGLVLTLDPEVRIAPAESSVAVRADGTLYLT 1817
Oy 1287 -----GAG----- 1289
Db 1818 GGLGLGLRVAWGLAERGAQGLVJVGSGAASAEGRAAVALLEAHGARVTYAKADVADS 1877
Oy 1290 ----- 1289

```

```

Db 1878 QIERVLEVTASGMPLRGVVAAGLVDDGLMOQTPARFRTWGPVKVQALHHTLTREA 1937
Oy 1290 ----- 1289
Db 1938 PLSTFVLYASAGLFGSPGCGNTAAANAAPLDALSHHRAQGLPALSTIDKMFTEVMAVA 1997
Oy 1290 -----PST----- 1292
Db 1998 QENRGARQISGMKRGITPDEGLSALRLLEGDRVQGVIPITPQWVEFYATASRLS 2057
Oy 1293 -----GDDLDLRLASAPARAALAEFLRTVOVQVLRTEIKYGAELF 1338
Db 2058 RLVTQRAVDRAQADRLLEQLASAPASAGLQDVVRVQSVLRLEPDKLEVAAPL 2117
Oy 1339 TRLGMDSLMAVELNRIRIASIKLSTTFLSTSPNIALAONLIL-DALATAL 1389
Db 2118 SSMGMDSLMSELNRIRIEALGVAAPALGWTTFYVAALTRWLDLALVRL 2169

RESULT 13
US-09-568-472-7
; Sequence 7, Application US/09568472
; Patent No. 638719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyt, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568, 472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-7

Query Match 57.4%; Score 4137; DB 4; Length 2439;
Best Local Similarity 43.4%; Pred. No. 8.5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

Oy 6 IERAABPIAIVGASCRLLPGGVLDLSGFMTLLSGSRDVTGVRPARMDAAMPDPDPAP 65
Db 26 LEQERTPIAIVGICGRFPGGADTPBAFWELDGRDAV--QPLDRMALVGVHPSSEVP 83
Oy 66 GKTPVTASPLSD- VACFDASFFGISPREALRMDPAHRLLEVCWEALENNAIIPSAVLG 124
Db 84 -----RWAGLITAEAVDFDAAFGTSPREARSILDPQORLLEVTWEGLEDAIGIAPQSIDG 138
Oy 125 TENGVFICIGISEYEALPQATASAEIDAAGGCTMSVAGRISYALGRCGPVADTA 184
Db 139 SRTGVFLGACSDSHVTAQORRE-EGDAVDITGNTLSVAAGRLSYTLGQPCLTVDTA 197
Oy 185 YSSSLVAVHLACQSLRSGECSTALAGVSLMSPSTLWMLSKTRALARDGCKAFSAEAD 244
Db 198 CSSSLVAIHLACSLRLARRESLDALAGVNMLLSKTMIAGRIQALPDPGCRFPDASAN 257
Oy 245 GFGRCGCAVVLKRLSGAPADGDRILAVIRGSAINHDASSGLITVNGSSQEIYLRAL 304
Db 258 GFVFGECGAVVLKRLSDAQRHGDRIWALIRGSAMNODGSRGTGIMADNVLAQEAALLREAL 317
Oy 305 ADAGCAASVGVYAHGTGTTIGDPIETIOALNAYYGRVAVPRLIGSVYTNLGHPEYA 364
Db 318 QSAKVDDGALIGVETHTGTGSLGPPIEVALRAVLGPARADGSCVCAKYTNLGHLEGA 377
Oy 365 SGITGLKVLVLSIOHQI PAHLHAQALNPRISWGLRLTVTRATPWPDMNTPRAGVSS 424

```

```

Db      378 AGVACILIKALALHHELIIPRNHFTLNRIIEGALALALATEPWPWPAAGPRFAGVSA 437
Qy      425 FGMSTNAHVLEAPATCTPAPERPAELIVLSARTASALDAQARLRDLETPSOC 484
Db      438 FGLSTGNVHVLEAPATVLPATPGRSALLVLSAKSMAALDAQARLSAHIAVPEQG 497
Qy      445 LGDVAFSLATTSAMEHRLAVALATSRREGALDLDAACQOTSPGAVRSIADSRKRLAVL 544
Db      498 LGDVAFSLVSTRSPMEHRLAVALATSRREALSALVAAQOTSPAGAARGAASPGKLAFL 557
Qy      545 FTGQAGQTLGMRGLYDVWASFRBAFDLCVRLFNOELDRPREVMMWABASPDALLDOT 604
Db      558 FAGQAGVPGMRGMLBEMAPAFRETFDRCVTLFDEHLHPLCEVMMWABESSRSLDOT 617
Qy      605 AFTQPALFTEFYALALMWSGVEPELVAGHSIGELVAACVAGVSLBEDAVFLVARGRL 664
Db      618 AFTQPALFTEFYALALFMSGVEPELVAGHSIGELVAACVAGVSLBEDAVFLVARGRL 677
Qy      665 MOALPAGGAMVSIAPADVAALVAPHAASVSIIVANAPDOVVIAGAGPVAIAAAMA 724
Db      678 MOALPAGGAMVSIAPADVAALVAPHAALVSIIVANPEOVVIAGAERFVQOIAAFPA 737
Qy      725 RGARTKALVSHAFSPILMAFMLBAFRVAESVSRPSIVLVSNLSGKACTDEVSPPCY 784
Db      738 RGARTKALVSHAFSPILMDPMLBAPFRVTSVTRRBSIALVSNLSGKCPCTDEVSAGCY 797
Qy      785 WVRHAREVVRPADGVKALHAAGAGTFVEVGPKSTLLGLVAPACMPDAPALLASSRAGDE 844
Db      798 WVRHAREVVRPADGVKALHAAGAGTFVEVGPKSTLLGLVAPACMPDAPALLASSRAGDE 857
Qy      845 PATVLEALGGLMANGVLVSMAGLFPSSGGRVPLPYPMQREBYVI---DTADDAAR-- 898
Db      858 AASALEALGGLFVWVGVSVTWVGVPSPGGRVPLPYPMQREBYVIAPVDRADGTGRAR 917
Qy      899 ----- 898
Db      918 AGCHPLLGEVFSVSTHAGRLMETLDRKRLPWGEHRAQGEVFRPGAGYLEMALSSGAE 977
Qy      899 ----- 904
Db      978 ILGDPIQVTDVLLIETLTFAGDTAVPVQVVTTEERPGRLRFQVAREBERGERARFRIHA 1037
Qy      905 ----- 904
Db      1038 RGVLRIGRVETPARSNLAALRLARLHAAPVAPAAITVGALEWGLQYGPALRGLAELEMRGEG 1097
Qy      905 ----- 904
Db      1098 EALGVRLEPEAGSATAYQVLMHPVLLDACVQMTVGAPADRDDEATPMAPEVGSVRLFORSP 1157
Qy      905 ----- 904
Db      1158 GELMCHARVSDQOASRWSADPELMDGTGAVVAISRLVVERLSAGVRRRADDWFILE 1217
Qy      905 ----- 908
Db      1218 LDWEPALGGPKITAGRWLLLGEGGGLGRSLCSALKAGHVVAAGDDTSTAGKRALIA 1277
Qy      909 ----- 908
Db      1278 NAFDQAPTAVVVHSSLDGGGLGPGLAGAGLADPRSPDVADALLESALMGCSVSL 1337
Qy      909 ----- 908
Db      1338 VQALVGMULRNAPRLMLTRGAQAAAAGDVSVVQAPLLGLGRTIALHEALRCSIVDLP 1397
Qy      909 ----- 952
Db      1398 AEPSEBADALLAELLADDAEEVALRGGRVLRVHRIPDQOREKVEPADDRPRELEI 1457
Qy      953 DEPGVLDHLVLRVTERRAFGLEVEIIVADAAGLSFNDVOLATGAMPDDLPGRKPNPLLLG 1012

```

```

Db      1458 DEPGALDQVLVRATGRARPGGEVEISVEAGLDSIDIQIALGVAPNDLPGEIEPLVIG 1517
Qy      1013 GECAGRIYAVEGNGVGLVGGPVIALNSGATATVTTSTAAVLVRRPOLSLAEAAAVVA 1072
Db      1518 SECAGRIYAVEGNGVGLVGGPVIALAGVATVTTSTATVLPPLGLSATEAAVLA 1577
Qy      1073 YLTAVYALDRILARLOPGERVLIHAATGGVAAVQMAQVGAEBVATAGTEKRAYLESL 1132
Db      1578 YLTAVYALDRILARLOPGERVLIHAAGVGLCAVRMAQVGAEBVATADTEENRAYLESL 1637
Qy      1133 GVRVYSDRSRPFVADVAPMTGGEVDVVLNSLSELIDKSFNLSHRGFEVELEKRDY 1192
Db      1638 GVRVYSDRSRPFVADVAPMTGGEVDVVLNSLSELIDKSFNLSHRGFEVELEKRDY 1697
Qy      1193 ADNOLGLRPFIRNLSFSLVDLRGMLEPAPVRRALBEELJGLIAGV----- 1239
Db      1698 ADTQGLPPLLRNLSFSSQVDRGMMLDQPARIRALLDELFGIVAGAISPUGSLRVGS 1757
Qy      1240 FTPLPILATLPIARVADAFRMAQOHLGKLVLTIGDPEVOIRIPHA----- 1286
Db      1758 LTPPEVETFPISRABAEFRRAVQOHLGKLVLTLDDEVRIRABAESSVAVRADGTIYLT 1817
Qy      1287 ----- 1289
Db      1818 GGLGGLGRVAGWLABRGAGQLVYGRGASASQRAAVALBAHARVTVAKADVADRS 1877
Qy      1290 ----- 1289
Db      1878 QIERVLEVTASGMPLRGVHAAGLVDDGLMOQTPARFRTWGPVKOGALHLTLTRE 1937
Qy      1290 ----- 1289
Db      1938 PLSPFVLYASAAGLFGSGGQNYAANAFLDALSHHRAQGLPALSDIMGMFTVGMAYA 1997
Qy      1290 ----- 1292
Db      1998 QENRGARQISRGMGITPDEGLSALARLLBGDRVQGVIPITPQWVEFYDATABRSL 2057
Qy      1293 ----- 1338
Db      2058 RLVTQRAVADRAGDRDLLEQLASAEPSAAGLQDVVRQVSHVNLPRDKLEVDAPL 2117
Qy      1339 TRLDMSLMAVELNRRIEASLKLKLTSTFSPNIALAONLL-DALATAL 1389
Db      2118 SSMGMSLSLELRRIEALGVAAAPALGMITYFVAIITWLLDDALVRL 2169

```

```

RESULT 14
US-09-567-899-7
; Sequence 7, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567, 899
; PRIOR FILING DATE: 2000-05-10
; PRIOR FILING DATE: 1999-06-17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2439
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-567-899-7
Query Match      57.4%; Score 4137; DB 4; Length 2439;

```


Best Local Similarity 43.4%; Pred. No. 8.5e-297;
Matches 933; Conservative 135; Mismatches 308; Indels 776; Gaps 13;

Qy	6	1EBAEDD1ATAYGASCRLLPGCVI	1DLSGFMTLL	EGSRDTPYGR	PAEEMDAANFDP	PDAP	65
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	26	LEQRTPE1ATVIGICRFPFGADTPEA	FELLDSGD	AV - -OPLDR	MALVGVH	PEEVEP	83
Qy	66	GKTPYRASFLSD - VACFDAS	FFG1SPR	ALBMDP	PAHRLLLEVC	EMALENMAIAP	124
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	84	----- RMAGLJLEA	NDGFDPA	FFG1SPR	ARBLDPQ	RLLEVTWEGLED	138
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	125	TETGVFIGI	GBSEPAAL	PQATASAE	IDAHGGL	GTWPSV	184
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	139	SRTGVFLG	ACSSDSDHTVA	AOQRR - EOD	AYDITGNTL	SVAA	197
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	185	YSSSLVA	HLACOSLRSG	CSFTALG	GVSLMSP	STVLWLS	244
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	198	CSSLVA	HLACRSRARS	SDALAG	VMMLSK	TMWLGRIQ	257
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	245	GFGRGEG	CAVVYLKRLSG	ARADDER	ILAVTR	SAINHD	304
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	258	GFVRGEG	CGMYLKR	LSDAQ	GRDRI	IMALIR	317
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	305	ADAG	CAASVGVYEA	HGTTLG	DP1E	IOALNA	364
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	318	QSA	RVDAGAI	GVETHE	TGTSLG	DP1E	377
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	365	SGIT	LLKVLVSL	IOGQIP	PAHLLA	QALN	424
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	378	AGV	AGLIC	IOALAL	HLHEL	IPRNLH	437
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	425	FGMSG	TNAHVLE	EARPAAT	CTP	PAPER	484
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	438	FGL	STGNHVVLE	EARPAAT	GPAT	GRS	497
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	485	LG	DAVFL	ATRS	AMEHRL	AVAA	544
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	498	LG	DAVFL	ATRS	AMEHRL	AVAA	557
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	545	FTG	GAQTL	GMGR	GLYDV	WAS	604
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	558	FAG	GAQY	PGMGR	GLME	AMP	617
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	605	A	FTOP	AL	TFE	YAL	664
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	618	A	FTOP	AL	TFE	YAL	677
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	665	MO	AL	PAG	AMV	SI	724
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	678	MO	AL	PAG	AMV	SI	737
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	725	RG	ART	KL	HSV	SHAF	784
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	738	RG	ART	KL	HSV	SHAF	797
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	785	WVR	HA	REV	VR	PAD	844
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	798	WVR	HA	REV	VR	PAD	857
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	845	PAT	VE	AL	AG	LMA	898
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	858	AA	S	AL	EA	L	917
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Qy	899	---	---	---	---	---	898
		: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	918	AG	H	P	L	E	977
		: : : : :	: : : : :	: : : : :	:		

Dd	1038	KGVLRIGRVETTPAKSNLIALRLARLHAAVPAAALYCALAEMGLOYPALGLAELMRGEG	1097
Qy	905	-----	904
Dd	1098	EALGRVRLPEAAGSATYQHLPHVLDDACVQMIVGAFADRDPEATPMVAVEVGSVRLFORSP	1157
Qy	905	-----	904
Dd	1158	GELWCHARVVSDDQOASSRWSADPELMDTGCAVVAEISRLVVERTLAGVRRDADWFLF	1217
Qy	905	-----GAGH-----	908
Dd	1218	LDMPEALGPKITAGRWMLLLEGCGGLGRSLCSALKAKAGHVVHAAGDSTSTAGMRALLA	1277
Qy	909	-----	908
Dd	1278	NAFPGQAPTAVNHLSLDDGGQQLGPGAGALDAPRSDDVDADLESALMRGDSVLSL	1337
Qy	909	-----	908
Dd	1338	VQALVGMDELNRAPRLMLTRGAQAAAGDVSVVQAPLLGKRTALBHAELRCISVLDP	1397
Qy	909	-----DEVEEGAVRGGDRSRALDHPPESSGGRKVEYEAAGDPFLEI	952
Dd	1398	AEPEGEADALLAELLADABEEVAALRGGDRLVALVRLPDQORREKVEEPAGDRPFREI	1457
Qy	953	DEPVLPHLVLRTERRAPGIGVEYEVAVDAAGLSFNDVQALAGVPPDLPGKPPRPLL	1012
Dd	1458	DEPALLQVLURATNGRAAPGEGVEISEVAAGLDSIDIQALGAPAPDLPGOEBLEPLVLG	1517
Qy	1013	GECAGRIVAVEGNGVLVGGPVIATLSAGAFATHVTTSAALVLEPPOALSIEAAPVA	1072
Dd	1518	SECAGRIVAVEGNGVLVGGPVIATLAAGVAFATHVTTSATVLEPRPLGSTAEAAAPLA	1577
Qy	1073	YLTWYVALDRARLOPBERVLIHAATGCVGLAAVQMAQNYGAEFHAATNGPBERKAYLESL	1132
Dd	1578	YLTWYVALDRARLOPBERVLIHAAGVGGLCAVRMAQVRGAEVYATADPENNAAYLESL	1637
Qy	1133	GVRVVSRSRSRPFADVRAMTGGEGVNVVNLNSLSELIDKSFNLRSGHGFVLEKGRDCY	1192
Dd	1638	GVRVVSRSRSRPFADVRAMTGGEGVNVVNLNSLSELIDKSLNVLKRCGLVKLGRDDC	1697
Qy	1193	ADNOLGLRPFLRNLSFSLVDLRCGMLEBRPARVRALTEELGLINAGV-----	1239
Dd	1698	ADNOLGLRPFLRNLSFSGOVOLDRGMLDQPARIRALDELFGVLAAGAISPLGSLGRVGS	1757
Qy	1240	FTPPPIATLPIARVADAFRSMAOQHLGKLVLTLDGPEVOIRIFTHA-----	1286
Dd	1758	LTPPPEVTFPISRAAEFRFMAOQHGLKLVLTLDPEDEVRIRAPAESSVAVRADGYLVLT	1817
Qy	1287	-----GAG-----	1289
Dd	1818	GGLOGGLGRVAGWLAERGAQOLVLRSGSAASAEQRAVAALAEHNGARVTYAKADVDRS	1877
Qy	1290	-----	1289
Dd	1878	QIERVLAENVASGMPLRGVHAAGLVDDGLMQOTPARPFTVMGPKXGAGALHHTLTREA	1937
Qy	1290	-----	1289
Dd	1938	PLSEFVLVYASAGLFGSPGOQNYAANAFLDALSHHRAAGLPALSIDMGVTEGVNAVA	1997
Qy	1290	-----PST-----	1292
Dd	1998	QENRGARQISRGNRGITPPDEGLSALAARLEBGDRVQTCVITPTTPQWNEFPATASRRLS	2057
Qy	1293	-----GDRDLDRIRLASAAPARAARAALEAFRTOVSQVLRTEIKVGAELF	1338
Dd	2058	RLVTTQRAVADRTAGDRDLEQLASAPSAAPAGLQOVNAVQVSHVLRLEPDKIEVDAPL	2117
Qy	1339	TRLOMSLMAVELRNRIEASIKLKLSTTFPISTSNIALLAONLL-DLALATL	1389
Dd	2118	SSKMDISLMSLELRNRIEAALGVAAAPALGMVTFVTAIRFMWLDLVLVRL	2169

RESULT 15
US-09-413-814-70

Sequence 70, Application US/09413814

Patent No. 6225064
GENERAL INFORMATION:

GENERAL INFORMATION

ABPLT.CANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, Co.

APPLICANT: Beyer, Stefan

APPLICANT: Bloecker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Goldberger, Steven

APPLICANT: Hoffle, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Har

TITLE OF INVENTION: DNA seq

TITLE OF INVENTION: heteropolymer

FILE REFERENCE: PCT/US 99/23535

CURRENT FILING DATE: 1989-10-07

EAB/ITER APPLICATION NUMBER: DE 198 46 493

EARLIER FILING DATE: 1998-10-09

NUMBER OF SEQ ID NOS: 107

; SOFTWARE: PatentIn Ver. 2.1.

; SEQ ID NO 70

; LENGTH: 22

```

; TYPE: PRT
ODCINTOM: 0

```

ORGANISM: SC
HS-09-413-814-70[illegible]

query match

Best Local Similarity 43.2%; Pred. No. 3e-274;

Matches 859; Conservative 118; Mismatches 245; Indels 767; Gaps 9,

Qy	168	ISVALLRBPCEVADPVAYSSSLVAHMLACQSLRSCECTALAGGSLMLSPSTLWMLSKT	22.7
Qy	1	MSVTLGLQGPCLTVDVYCSSSLVAHMLACRSLRAESPDLLAGCNMLLSKTMIMLGR	60
Qy	228	RALARPGRCKAPSAEADGFRGEGCAVYVYKRLSGARDGRILLAVIRGSAIINHGAASSG	28.7
Db	61	QALSPDGHCRITDASANGFVRGEGGMVYVYKRLSPDQHGRRIVALIRGSSIMNDGSGT	120
Qy	288	LTVPNSSSOEIVYKRLADAGCAASSVGEVYHGGTTLGDPPIEIOALNAVYGLGRVAT	34.7
Db	121	LMAPNVLAGEALLRELQASRYDAAGIGVEYHGGTSLGDPPIEVALRAVYLGARADGS	100
Qy	348	PLIISGVKTNLGHPEYASGITTELLKVYVLSLOHGQIPAHLLHQAOLNPRI SWGDLRIVTRA	40.7
Db	181	RCVLGAVKTNLGHLEBAMAGVAGLIRAAALAHNELPRMIAHPTTNIPRIRIEGTALALATE	24.0
Qy	408	RTPVPMNTPRRAGVSSFCMGSTNAHVYLEAPATCTPPAPERPAELLYVSTARTASALD	46.6
Db	241	PVWPMPAGRPFRAGVSAFELSGSTNVHYLEAPAVLAPATPGSSABELLYVSASSAALD	30.0
Qy	468	AQAPLRDLHLETTYPSQCLGDVAFSLATTRSAMERFLAVAAATSRBGLRAALDAAAGQSTP	52.7
Db	301	AQAPLRISAIIAAVPEEGCLGDVAFSLVSTRSPMEHRLAAATSRREALTSALTEVAAGQSTPA	36.0
Qy	528	GAVRSLADSSRGCLAF.LFTGGGAQTLQMGGRGLYDVWMSFREAFPLCYRLPMQELDRPLRE	58.7
Db	361	GAARGRAAASPGLAF.LFPGGAQVPGMGGRGLMEAMPFRFETFDRCVTLVFRRLHQLCE	42.0
Qy	588	VMAAPASVDAALDDTAFTOPALFTFEYALAAALWRSGVPELVAGHSIGELVAAACVAG	64.7
Db	421	VMAAPGSSRSSSLDDTAFTOPALFALFYALALFRSGVPEPELVAGHSIGELVAAACVAG	48.0
Qy	648	VSELAEAVFLVARGGLMQLPAGGAMVSIAPAEADVAAAVAPHAASVIAAVNAPPOV	70.7
Db	481	VSELAEADVIRLVARGGLMQLPAGGAMVSIAPAEADVAAAVAPHAALVIAAVNAPPOV	54.0

QY	708	IAGGQGVVHAIAAAMARGARTKLAHSHAFSLPMAFMAEAFRVAESVYRPSITLV	767
Db	541	IAGKEVVOQIAAFAARGARTKPLHSHAFSLMDPMLAEFRVRESYTYRRPSITLV	600
QY	768	SNLSGKACTDEVSPPGYVWHAREVVFADGVKLAHAAGTFVEVPSKTLGLVPACM	827
Db	601	SNLSGKCTDEVSAPGYVWHAREVVFADGVKLAHAAGTLFVEVGPKPLTGLVPACL	660
QY	828	PDARFALLASSRAGRDPAVYLEAGLMAVGLVSNAGLPFSGSGRVPLPTYWQERY	887
Db	661	PDARFVLLPASRARDEMAALFALGFMVWGSVTWSGVFSSGGRVPLPTYWQERY	720
QY	888	WI-----DTKADDAAR-----	898
Db	721	WIEAPVUREADGTGRARAAGHPLLGEVFSVSTHAGLRMTETTLDRKRLPMLGHRAGGV	780
QY	899	-----	898
Db	781	VFPAGYLEMALSSGAEILGDGPQVTDVVLITFLTPAGDTAVPVQVVTTEERPGRLRFQ	840
QY	899	-----GDRAP-----	904
Db	841	VASREPERRAPFRIHARGVLRIGRVETPARSNLALRLARLHAAPVAAAITYGALAEKGL	900
QY	905	-----	904
Db	901	QYGFALRGLABLWRGEBEALGRVRLPEAAGSATAYQLHPVLIDACVOMITYGAFADREBAT	960
QY	905	-----	904
Db	961	PMAPVEGVSVLRPQRSPEGLMCHARVVSDDQOASSRKSADPELMDGTGAVAEISRLVE	1020
QY	905	-----GAGH-----	908
Db	1021	RLASGVRRDADWFLFELDEPAPALGGPKITAGRWLLTGEGGGLRSLCSALKAAGVWV	1080
QY	909	-----	908
Db	1081	HAAGDDTSTAGMRLLANAFDGAFTAVVHLSTLDDGGQLGPGLAGOALDAPRSPDVDA	1140
QY	909	-----	908
Db	1141	DALESALMRGDSVLSVLQALVCGMDLNNARLMLTLTGQAAAAAGVSVVQAPLLGIRT	1200
QY	909	-----DEVEGAVRGDRRSARLDHPPEBSG	935
Db	1201	IATLHAELRCISVDLPABEPEGADALLAEHLADDAEEVVALRGDRLVARLVHRLPDAQ	1260
QY	936	RREKVEAAGRPPLPLETIDEBGVLDHLVLRTERAPBLGVEYETAVDAAGLSFNDVOALG	995
Db	1261	RREKVEAAGRPPLPLETIDEBGALDQVLRATGRAPBPGEVEISVEAAGLDSIDIQALG	1320
QY	996	MVPDDLPGKPNPPLLGGECAGRIVAVGEGVGLVWGPYIALSAGAFATHTVTSAAVL	1055
Db	1321	VAPNDLGEBELEBPLVJGSECAGRIVAVGEGVGLVWGPYIALAAGFAHTVTSATLV	1380
QY	1056	PRPOLASALEAAMPVAYLTAMVALDRIARLOPGERVLIHAATGSGVGLAAVQAQHVGAE	1115
Db	1381	PRPLGLSATEAAMPVAYLTAMVALDKVAHLOAGERVLIHAAGVGGLCAVRMAQRGAEE	1440
QY	1116	VHATATGPEKRAYLESIGRVYSSRSRDRPADVRAATGEGVGVUNVLSLGEILDSFN	1175
Db	1441	VYAADNPENRAYLESIGRVYSSRSRGRFTVDHAATDEGVGVUNVLDLSGGEIDKSLM	1500
QY	1176	LIRSHGFVLEKGRDCVADNOLGLRPFLRNLISFLVLRGMLERPARVRLAEELGLI	1235
Db	1501	VLRQCGRLVLRGRDDCADTOPGLPPLLRNFSSQVLRGMLDQPARIRRLDELGLV	1560
QY	1236	AAGV-----FTPEPITLPIARVADAFRSMAQOHLKVLTLTGDBEVOIRI	1282
Db	1561	AAGHISPLGSGLARVGSUTPRPVETPFISRAABAFRMAQOHLKVLTLTDDEVRIRA	1620
QY	1283	PTHA-----GAG-----	1289

```
Db 1621 PAESSVAVRADGYLVYTGGLGLGRVAGWLAERAGQLVWGRSGAASAEORAAVALE 1680
QY 1290 ----- 1289
Db 1681 AHGARVTVAKADVADRSQIERVLRBTASGMPLRGVVAAGLVDDGLMOQTPARFTVM 1740
QY 1290 ----- 1289
Db 1741 GPKVQALHHTLTREAPLSPFVLYASAAGLFGSPGQGNVAANAFLDALSHHRAQGLP 1800
QY 1290 ----- 1289
Db 1801 ALSIDMGWTEVGMAVAQENRGARQISKMRGITPDEGLSALARLLEGDRVQGVIPITP 1860
QY 1290 ----- 1321
Db 1861 RQWVEFPATPASRLSLVTTORAVADRTAGDRDLLEQLASAEPSARAGLLODVVRQV 1920
QY 1322 SOVLRTPEIKVGAELFTRLGMDSIMAVELNRRLEASIKLSTTFSTSPNIALAQL 1381
Db 1921 SHVLRLEPEDKIEVDAPLSSMGMDSIMSELNRRTEAALGVAAAPALGWTYFTVAATRWL 1980
QY 1382 L-DALATL 1389
Db 1981 LDDALVVRL 1989
```

Search completed: October 2, 2003, 17:35:52
Job time : 71.7527 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 95.763 Seconds
(without alignments)
2355.298 Million cell updates/sec

Title: US-10-014-717-2

Perfect score: 7210
Sequence: 1 VADPPIRAEDPIAIVGAS.....GVQNDPWSGADQDWEIHAL 1421

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneeq/geneeq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7210	100.0	1421	21	AA158573
2	4137	57.4	2439	21	AA158578
3	3311.5	45.9	7257	21	AA158576
4	2531.5	35.1	2153	22	AA183974
5	2363	32.8	4038	22	ABG9871
6	2355.5	32.7	5435	22	AAE10145
7	2351	32.6	4106	22	ABG9872
8	2331.5	32.3	11096	22	AAE10129
9	2319.5	32.2	2149	24	ABP57679
					Sorangium cellulos
					Sorangium cellulos
					Sorangium cellulos
					Amino acid sequenc
					S. cinnamomensis M
					Streptomyces nouris
					S. cinnamomensis M
					Streptomyces nouris
					Saccharopolyspora

10	2273	31.5	2152	20	AA19298
11	2273	31.5	2152	22	AA1870966
12	2260.5	31.4	7068	22	AAE10142
13	2260.5	31.4	9477	22	AAE10144
14	2250	31.2	6095	22	AAU10701
15	2245.5	31.1	3798	21	AA158577
16	2194	30.4	5087	22	AAU10700
17	2165	30.0	6797	22	AA1831558
18	2150	29.8	1841	18	AA1822605
19	2135	29.6	4132	22	ABG9870
20	2128	29.5	3816	21	AA192708
21	2127	29.5	3170	20	AA192929
22	2127	29.5	3170	22	AA1870967
23	2125	29.4	4630	21	AA177177
24	2119	29.4	4933	24	ABP57681
25	2113	29.3	4924	22	AA193930
26	2113	29.3	4928	20	AA193930
27	2108.5	29.2	3562	22	AA182231
28	2107.5	29.2	3567	14	AA184431
29	2104	29.2	2238	22	ABG9869
30	2103	29.2	4630	18	AA19629
31	2097.5	29.1	1366	22	AA1810127
32	2097	29.1	3167	24	ABP57680
33	2091.5	29.0	5588	20	AA19301
34	2091.5	29.0	5588	22	AA19301
35	2078	28.8	4344	24	ABP57678
36	2065.5	28.6	5564	24	ABP57682
37	2063.5	28.6	3413	19	AA182849
38	2052.5	28.5	3739	21	AA177193
39	2052.5	28.5	3739	21	AA177201
40	2052.5	28.5	12199	21	AA177180
41	2051.5	28.5	4551	21	AA18637
42	2051.5	28.5	4551	21	AA18637
43	2051.5	28.5	4551	21	AA18637
44	2051.5	28.5	4613	21	AA177192
45	2051.5	28.5	4613	21	AA177200

ALIGNMENTS

RESULT 1	AA158573	standard; Protein; 1421 AA.
XX	AA158573;	
XX	AC	
DT	10-APR-2000	(first entry)
DE	Sorangium cellulosum type I polyketide synthase EPOS A.	
KW	EPOS A, type I polyketide synthase; epothione biosynthesis; thiazole ring formation; taxol substructure; anticancer.	
OS	Sorangium cellulosum	
XX	OS	
XX	PN	MO9966028-A2.
PD	23-DEC-1999.	
XX	PF	16-JUN-1999; 99WO-EP04171.
XX	PR	18-JUN-1998; 98US-0099504.
XX	PR	24-SEP-1998; 98US-0101631.
XX	PR	05-FEB-1999; 99US-0118906.
XX	PA	(NOVS) NOVARTIS AG.
XX	PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GBS MBH.
PI	Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;	
DR	WPI; 2000-097741/08.	
DR	N-PSDB; AA255887.	

SpnB a polyketide
S. spinosa protein
Streptomyces nours
Streptomyces nours
Amino acid sequenc
Sorangium cellulos
Amino acid sequenc
Pimaricin biosynth
Ty lactone synthase
S. cinamomensis M
S. antibioticus 8,
SpnB a polyketide
S. spinosa protein
S. venezuelae vep
Saccharopolyspora
S. spinosa protein
SpnB a polyketide
Polyketide synthas
eryA region polype
S. cinamomensis M
Streptomyces venez
Streptomyces nours
Saccharopolyspora
SpnB a polyketide
S. spinosa protein
Saccharopolyspora
S. spinosa protein
Saccharopolyspora
A. mediterranei ri
S. venezuelae macr
S. venezuelae pik
Amino acid sequenc
Narbonolide syntha
S. venezuelae narb
S. venezuelae macr
S. venezuelae pik

XX New isolated epothilone synthase genes, used for the recombinant
 PT production of epothilone for use in cancer therapy

XX Claim 12; Page 105-109; 174pp; English.

CC This sequence represents a Sorangium cellulosum type I polyketide
 CC synthase, EPOS A, which is one of several epothilone biosynthetic
 CC enzymes encoded by a 68.75 kb contig. Epothilones A and B are
 CC 16-membered macrocyclic polyketides with an acylcysteine-derived
 CC starter unit; polyketides being synthesised from two-carbon building
 CC blocks, the beta-carbon of which always carries a keto group. Each round
 CC of two-carbon addition is carried out by a complex of enzymes known as
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
 CC EPOS A (AA158573) and EPOS P (AA158574) are involved in formation of
 CC the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
 CC and EPOS E (AA158575-158578) are involved in polyketide backbone
 CC formation. EPO F (AA158579) is an epothilone macrolactone oxidase, and
 CC the proteins Orf 3 (AA158582) and Orf14 (AA158593) are thought to be
 CC involved in transport. Epothilones mimic the biological activity of
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic
 CC compositions. Epothilones exhibit a much lower drop in potency against a
 CC multiply drug-resistant cell line compared with taxol, and are
 CC considerably less efficiently exported from such cells by the multidrug
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
 CC epothilones as anticancer agents, they are problematical to produce on a
 CC large scale. Epothilones are too complex for industrial scale chemical
 CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
 CC poor yields of epothilones. The nucleic acids of the invention may be
 CC used for the recombinant production of epothilones in a heterologous host
 CC that is more amenable to fermentation.

XX Sequence 1421 AA;

Query Match 100.0%; Score 7210; DB 21; Length 1421;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VADRIPIERAADPIAIVGASCRIPGVIDLSGFWTLLEGSDRTGVRPERMDAAAFPP 60
 DB 1 VADRIPIERAADPIAIVGASCRIPGVIDLSGFWTLLEGSDRTGVRPERMDAAAFPP 60
 QY 61 DDDAGKPTVTASFLSDVACFDASPFGISPREALRMDPAHRLLEVCHEALLENAAIAS 120
 DB 61 DDDAGKPTVTASFLSDVACFDASPFGISPREALRMDPAHRLLEVCHEALLENAAIAS 120
 QY 121 ALVGETGVFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGLRCPVA 180
 DB 121 ALVGETGVFIGIGSEYEALPQATASAEIDAHGGLGTMPSVGARISYALGLRCPVA 180
 QY 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMSPSTLYWLSTKRALADGRCKAFS 240
 DB 181 VDTAVSSSLVAVHLACQSLRSGECSTALAGVSLMSPSTLYWLSTKRALADGRCKAFS 240
 QY 241 AADAGRGREGGCAVYVLRKLSGRADGRIILAVIRGSAINHDPGASGLTPVSSGSEIYL 300
 DB 241 AADAGRGREGGCAVYVLRKLSGRADGRIILAVIRGSAINHDPGASGLTPVSSGSEIYL 300
 QY 301 KRALADAGCAASSVGVEAHGTTGLDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
 DB 301 KRALADAGCAASSVGVEAHGTTGLDPIEIQALNAVYGLGRDVATPLLIGSVKTNLGH 360
 QY 361 PEYASGITGLKVVYLSLQHGQIPAHILHAQALNPRISWGDLRLTVTRARTPMPDWTPRRA 420
 DB 361 PEYASGITGLKVVYLSLQHGQIPAHILHAQALNPRISWGDLRLTVTRARTPMPDWTPRRA 420
 QY 421 GVSSTGMSGTNAHVLEEARPAATCTPPAPERPELVLVSARTASALDAARLHOLETY 480
 DB 421 GVSSTGMSGTNAHVLEEARPAATCTPPAPERPELVLVSARTASALDAARLHOLETY 480
 QY 481 PSQCIQDVAFSLATTSAMEHRLAVALATSRREGLRALDAAAGQTSPGAIVSIADSSRCK 540
 DB 481 PSQCIQDVAFSLATTSAMEHRLAVALATSRREGLRALDAAAGQTSPGAIVSIADSSRCK 540

QY 541 LAFLFTGOGAQTLLGNGRLYDVWSAFREAFDLCVRFNQEILDRPLREVMMAEPASVDAL 600
 DB 541 LAFLFTGOGAQTLLGNGRLYDVWSAFREAFDLCVRFNQEILDRPLREVMMAEPASVDAL 600
 QY 601 LDQTAFTOPALFTPEYALAAALMRSGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 660
 DB 601 LDQTAFTOPALFTPEYALAAALMRSGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVA 660
 QY 661 RGRIMQALPAGAMVSIAPADVAANAAPHAASVIAAANAPODVITAGAGQPVHA1TA 720
 DB 661 RGRIMQALPAGAMVSIAPADVAANAAPHAASVIAAANAPODVITAGAGQPVHA1TA 720
 QY 721 AMAAGARTKALHVSAAHSPMLMAFPAFGRVAFSVSYRRPSTLYVNSLGSKACTDEVS 780
 DB 721 AMAAGARTKALHVSAAHSPMLMAFPAFGRVAFSVSYRRPSTLYVNSLGSKACTDEVS 780
 QY 781 SPGYVVRARAREVVRPADGVKALHAAGACTFVVGPKSTLLGLVPCMCMDARPALASSRA 840
 DB 781 SPGYVVRARAREVVRPADGVKALHAAGACTFVVGPKSTLLGLVPCMCMDARPALASSRA 840
 QY 841 GRDEPATVLEALGLMAVAGLVSWAGLFPSSGRRVPLPTYWQRRVYIDTKADDAARGD 900
 DB 841 GRDEPATVLEALGLMAVAGLVSWAGLFPSSGRRVPLPTYWQRRVYIDTKADDAARGD 900
 QY 901 RRAFGAGHDEVVEGAGVAGCGRRSARLDHPPESSGRREKVAAGDRPRLTIDEGVLDH 960
 DB 901 RRAFGAGHDEVVEGAGVAGCGRRSARLDHPPESSGRREKVAAGDRPRLTIDEGVLDH 960
 QY 961 LVLRTERARRAAGLSEVETAVDAAGLSFNDVOLALGMVDDLPKGNPPLLGGECAGRTV 1020
 DB 961 LVLRTERARRAAGLSEVETAVDAAGLSFNDVOLALGMVDDLPKGNPPLLGGECAGRTV 1020
 QY 1021 AVGEGVNLVWGQPVIALSAGAFATHTVTTSAALVPRPOALSIFAAMPVAYLTAMTAL 1080
 DB 1021 AVGEGVNLVWGQPVIALSAGAFATHTVTTSAALVPRPOALSIFAAMPVAYLTAMTAL 1080
 QY 1081 DRIARLOGERVLTJHAAGVGVGLAAVQWAQHVGAHVHATAGTPEKRAVLESIGVYVSDS 1140
 DB 1081 DRIARLOGERVLTJHAAGVGVGLAAVQWAQHVGAHVHATAGTPEKRAVLESIGVYVSDS 1140
 QY 1141 RSDRFVAADVAMTGGEGVDVVLNSLSGELIDKSFNILRSHGRFVELGRDCYADNQLGR 1200
 DB 1141 RSDRFVAADVAMTGGEGVDVVLNSLSGELIDKSFNILRSHGRFVELGRDCYADNQLGR 1200
 QY 1201 PFLRNLSFSLVDLRGMLEPRARVALLIEELGLTIAAGVFTPPPIATLPIARVADAFRSM 1260
 DB 1201 PFLRNLSFSLVDLRGMLEPRARVALLIEELGLTIAAGVFTPPPIATLPIARVADAFRSM 1260
 QY 1261 AQAQHLGKLVLTLDGPEVOIRIPTHAAGAPSTGDDLDRLASAPARAPAALEAFRTQ 1320
 DB 1261 AQAQHLGKLVLTLDGPEVOIRIPTHAAGAPSTGDDLDRLASAPARAPAALEAFRTQ 1320
 QY 1321 VSQVARTPEIKVGAELFTRLGMSLMAVELRNRIEASLKLKLTSTFTSTPNIALAQN 1380
 DB 1321 VSQVARTPEIKVGAELFTRLGMSLMAVELRNRIEASLKLKLTSTFTSTPNIALAQN 1380
 QY 1381 LLDALATLSLERRVAENLRAGVONDFVSSGADQDWEIITL 1421
 DB 1381 LLDALATLSLERRVAENLRAGVONDFVSSGADQDWEIITL 1421

RESULT 2

AA158578 standard; Protein; 2439 AA.

AA158578;

10-APR-2000 (first entry)

Sorangium cellulosum type I polyketide synthase EPOS E.

EPOS E; type I polyketide synthase; epothilone biosynthesis;


```

QY 909 ----- 908
Db 1338 VQALVGMRLRNA PRLMLLTRGAQAAAAGDVSVQAPLLGLGRTIALENALRCISVDLP 1397
QY 909 ----- DEVEEGGAVRGGRRSARLDHPPESSQRRREKVEAAGDRPRLRI 952
Db 1398 AEPGEADALAEALADDAEEVVALRGGRRLVAKRLVRLPDQORREKVEAAGDRPRLRI 1457
QY 953 DEPGVLDHLVLRVTERBARGLGEVEIADVAGLSFNDVOLALGMVDDLPKGNPRLILG 1012
Db 1458 DEPGALDQVLRLATGRRAPGGEVEISVBAAGLSDIDQLALGVANPDLPGSEIERLVLG 1517
QY 1013 GECAGRIYAVGSGVNCVVGQVIVIALSAGAFATHTTSAALVLRPQALSAIEAAMPRA 1072
Db 1518 SECAGRIYAVGSGVNCVVGQVIVIALSAGAFATHTTSAALVLRPQALSAIEAAMPRA 1577
QY 1073 YLTAWYALDRIRARLOPGEERVLJHAATGVLGAQVQAQVGAHVHATAGTPEKRAVLESL 1132
Db 1578 YLTAWYALDRIRARLOPGEERVLJHAATGVLGAQVQAQVGAHVHATAGTPEKRAVLESL 1637
QY 1133 GVRVYVSDSSDRFVADVRAWTGEGVDVVLNSLSEGLIDKSFNLRSHGRFVELGRDXY 1192
Db 1638 GVRVYVSDSSDRFVADVRAWTGEGVDVVLNSLSEGLIDKSFNLRSHGRFVELGRDXY 1697
QY 1193 ADNOLGRLPRLNLSFLVLAGMMLERARVRLLEELGLIAGV----- 1239
Db 1698 ADNOLGRLPRLNLSFLVLAGMMLERARVRLLEELGLIAGV----- 1757
QY 1240 FTTPPIATLPILARVADAFRMAQAQHLGLVLTIGDPVEQIRLPTHA----- 1286
Db 1758 FTTPPIATLPILARVADAFRMAQAQHLGLVLTIGDPVEQIRLPTHA----- 1817
QY 1287 -----GAG----- 1289
Db 1818 GGLGGLGLRLVAGMLAERGAQGLVLRSGSAGASAEQRAVAALBAHGARTVAKADVADR 1877
QY 1290 ----- 1289
Db 1878 QIERVIREVTASGMLRGVVAAGLVDDGLMQQTPARPTVMGPKVQALHLHTLTREA 1937
QY 1290 ----- 1289
Db 1938 PLSFVLYASAAGLFGSPGQGVYAAANALDLASHHRAQGLPALSIDKMTVEGMAYA 1997
QY 1290 ----- PST----- 1292
Db 1998 QENRGARQISRGMRGITPDEGLSALARLLEGDREVQGVIPITPQWVEFYPATASRRLS 2057
QY 1293 -----GDRDLDRLASAPARAALAEFLRTQVSQVLRPEIKVGAELF 1338
Db 2058 RLVTQRAVADRTAGDRDLLEQALSAEPARAGLQDVVRVSVHRLRPEKIEVDAPL 2117
QY 1339 TRLAGDSLMAVELARRNIEASLKLSTPLSTSPNIALIAQNL-DALATAL 1389
Db 2118 SSMGMDSLMSLELRNRIEALGVAAAPALGMWYPTVAATRLWLDALVRL 2169

```

```

PN MO9966028-A2.
XX 23-DEC-1999.
PD 16-JUN-1999; 99WO-EP04171.
XX 18-JUN-1998; 98US-0099504.
PR 24-SEP-1998; 98US-0101631.
PR 05-FEB-1999; 99US-0118906.
XX (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PI Schupp T, Liçon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
DR WPI; 2000-097741/08.
DR N-PSDB; AA255887.
PT New isolated epothonone synthase genes, used for the recombinant
PT production of epothonone for use in cancer therapy
XX Claim 12; Page 118-139; 174pp; English.
PS This sequence represents a Sorangium cellulosum type I polyketide
CC synthase, EPOS C, which is one of several epothonone biosynthetic
CC enzymes encoded by a 68.75 kb contig. Epothonones A and B are
CC 16-membered macrocyclic polyketides with an acylglycine-derived
CC starter unit; polyketides being synthesized from two-carbon building
CC blocks, the beta-carbon of which always carries a keto group. Each round
CC of two-carbon addition is carried out by a complex of enzymes known as
CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC EPOS A (AA558573) and EPOS P (AA558574) are involved in formation of
CC the thiazole ring formation of epothonones, and EPOS B, EPOS C, EPOS D
CC and EPOS E (AA558575-58578) are involved in polyketide backbone
CC formation. EPO F (AA558579) is an epothonone macrolactone oxidase, and
CC the proteins Orf 3 (AA558582) and Orf14 (AA558593) are thought to be
CC involved in transport. Epothonones mimic the biological activity of
CC taxol, and may be substituted for taxol in cancer chemotherapeutic
CC compositions. Epothonones exhibit a much lower drop in potency against a
CC multiply drug-resistant cell line compared with taxol, and are
CC considerably less efficiently exported from such cells by the multidrug
CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC epothonones as anticancer agents, they are problematical to produce on a
CC large scale. Epothonones are too complex for industrial scale chemical
CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC poor yields of epothonones. The nucleic acids of the invention may be
CC used for the recombinant production of epothonones in a heterologous host
CC that is more amenable to fermentation.
XX
SQ Sequence 7257 AA;
Query Match 45.9%; Score 3311.5; DB 21; Length 7257;
Best Local Similarity 41.1%; Pred. No. 2.3e-254;
Matches 784; Conservative 162; Mismatches 432; Indels 531; Gaps 25;
QY 9 AABDPIATVAGSCRLPGGVLDLGSFWTLLEBSRDTGVGVPARMDAAAMPDPDPAPKXT 68
Db 3021 ASDEPIATVAGACRPPGVEDLESTWQLAGGVVSAVPAADRMDAAWYDPDEIPRT 3080
QY 69 PVTASPLSDVACPDASFFGISPREALMDPAHRLLEVCWEALENAAIASALVGTETG 128
Db 3081 VYTKAFRLDQRLDQATFFRISPREAMSLDQOQLLEVSWEALIASALVGTETG 3140
QY 129 VFIIGSEYEYEAALPQATASAEIDAAGLGTMPYAGACRISYALGLRCPCAVADPTAYSS 188
Db 3141 VFVGAPNEYTYTORLRGTGDAAGLYGCTGMMLSTAGRLSFPLGHPPTLAMDPAACSS 3200
QY 189 LVAVHLACQSLRSGCSTALAGVSLMSPSTIWLSTRALARDGRKAFSAADGFR 248
Db 3201 LVALHLACQSLRLGCDQDALGVGVVLAPEFVLLSMRLASPDGRCKTTSADADGTAR 3260
QY 249 GEGCAVVLKRLSGARADGRILAVIRGSAINHDAASGLTVPNGSOGIEIVLKRALADAG 308

```

Db 3261 GEGCAVVVVKLRDQAQRADSLIALIRGSVANNHDSGSLTVNPGAOALLRQALSOAG 3320
 Qy 309 CAASVGVYEAHGTGTTLDGPRIEIOALNVAVYGLGRDVAFTPLIGSVKTNLGHPEYASGIT 368
 Db 3321 VSPVDVPEVECHGTALGDPRIEYQALSEVYGPBGSGRRPLVIGAKAKNVNHLHLEASGLA 3380
 Qy 369 GLIKVYLSLQGGQIPAHLLHAQALNPRISWGDLRLTVTRARTPWDPMTNTPRPAVSSFGMS 428
 Db 3381 SLIKVAVLALRHROIIPAQDELGNLPHMLPMTLPPVAVPRKAVPWGRCARPRRAGVSAFGLS 3440
 Qy 429 GTNAAVYLEAPATCTPAPERPAELVLSARTSALDQAARLDLETPYPCQCGDV 488
 Db 3441 GTNAAVYLEAPERPAERPAERPAERPAERPAERPAERPAERPAERPAERPAERPAERPA 3500
 Qy 489 AFSIATTSAMEHRLAVATSRREGRLAALDAAQOTSPCAVRSIADSGRKLAFLEFTGQ 548
 Db 3501 AFSIATTSRPMHRLAITSREALRGALDAAQOKTQCAVRCGAVSRKGLAFLEFTGQ 3560
 Qy 549 GAQTLGMRGLYDVMSAFREAFDLCVRLFNQELDRPREVMAABEASVDAALLDQTAFTQ 608
 Db 3561 GAQPMGMRGLYETMPAFREAFDRCVAFDREIDQPLREVMAAFAGLQAARLDQTAFAQ 3620
 Qy 609 PALTFEEVLAALMRSMGVEPELVAGHSIGELVAAVCAVGFSLDEAVFLVAARGLMQAL 668
 Db 3621 PALPALVLAALMRSMGVEPELVAGHSIGELVAAVCAVGFSLDEAVFLVAARGLMQAL 3680
 Qy 669 PAGGAMVSIABEADVAALVAAPHAASVSIAAVNAFDVVIAGAGPVHAIAMAAARGAR 728
 Db 3681 PAGGAMVAIAASEVAAVSAVPHATVSIIAVNGDAVVIAGAEVQVLAALGATRAARKIR 3740
 Qy 729 TKALVSHAFHSPLMAPMLEAFGRVAESVSYRPSIVLVSNLSGKACTDEVSPGYVVRH 788
 Db 3741 TKRLAVSHAFHSPLMDPMLDEFORVAATIAVRAADPRPVASVNTGVHAPRIATPEYVVRH 3800
 Qy 789 AREVVRPADGVKALHAAGAGTFVEVGPSTLLGLVACMPDAPRALLASSAGDERPTV 848
 Db 3801 VRSVVRFGDGKALHAAGAAATFVEVGPVLLGLPACLGADAVLVSLRADRSSECEV 3860
 Qy 849 LEALGGLAAVAGLVSMAGLPPSGGRRVPLPTYPMOREYVMDTKADA----- 896
 Db 3861 LAALGAMVYAKWGALDMKGVFPDGGARRVALPMYPMOREHMDLTPRSAAPAGIAGRMFLA 3920
 Qy 897 ----- 896
 Db 3921 GVGLCMRGAVLHVLISIGRHPFLGDHLVFGVVVPGAFHVAVLISIAERWPERATEL 3980
 Qy 897 ----- 902
 Db 3981 TGVEFLKAIAMEPQVEVLAHLTPPEAGDGVLELATLAAETERRWTTHARGVQPTD 4040
 Qy 903 -APGA----- 906
 Db 4041 GAFGALPRLVLEDBRAIOPLDPAFGLDLRLSAVRIGWGLMRLWLDGCVGDEASLTLVPT 4100
 Qy 907 -GHD-----EVEGGA-----VR-GGD 921
 Db 4101 YPNADHVAFLPHILLDNGFAVSLSTRSEPDGDPPLPFAVERVMMARVAGVRCGGV 4160
 Qy 922 RRSa-----RLDHP-- 932
 Db 4161 PRSQAFGVSVFLVDTEGEVVAEVEGFVCRAPREPVFLROESGASTALYRLDWPBEAFLP 4220
 Qy 933 -----BSGR 936
 Db 4221 DAPARIESWVVAAPGSEMAALATRLNRCVLAEPKGLAALAGVSPAGVITCMEAGA 4280
 Qy 937 REKVEAAGR----- 946
 Db 4281 HEBAPAAQVATBEGLSVQALRDBRAVRLMVVTWGAVAVEAGERVQVATAPVWGLGRTVM 4340
 Qy 947 ----- 946
 Db 4341 QERPELSCTLVLEBEADARSADVLNLELGRADDETQVAFSGRRVARLVKATPEGL 4400

Qy 947 -----PFRLEIDPGLVLDHLVLRTERRAFGCEVEIAVDAGLSFNDVQALGMVDD 1000
 Db 4401 LVPDASEVRLBAGGKGLDQRLAPQAARPAEPGEVEIKVTASLNFRTVLAVGM----- 4456
 Qy 1001 LPGKPNPPLLIGECACRIYAVEGVNGLVYQVYVIALSAGAFTHYTTSALVLRPQA 1060
 Db 4457 YPGDAGR---MGSCACAVATVAGGVVHVAVDVMTL---GTLHRFTVDRLVVRPAG 4511
 Qy 1061 LSAIEAAMPAVYLTAVYALDRIRARLQGERVLIHAATGGVGLAAVQAOHVAEYHATA 1120
 Db 4512 LTPAOAATVPAFLTAMLALHDGNLNRGERVLIHAAAGVGMAVQIARNTIGAEVFA 4571
 Qy 1121 GTPKRAYLSLGY--RYVSDRSRDEADVAVRAWTGEVGVVNLNSLSEGLIDXSFNLLR 1178
 Db 4572 -SPSKMAVQMGVPRTHIASRRLTEFAETFRQVTTGGRGVDVNLALAGEVDSLSLS 4630
 Qy 1179 SHGRFVELKRDCAVDNQ--LGLRPLRNTSFSVLVDLRCMMLERPARVALLBELGLIA 1236
 Db 4631 TGGRFLEMGKTD--TRDRAVAALAHFGRVYVFDILEL-----APDRTRILERVEGFA 4683
 Qy 1237 AGVFTPPPIATLPIARVADAFRSMAONHIGKVLTLGDPEVQIRIP-----HAGAP- 1290
 Db 4684 AGHURALPVAFAITKAEAFREMAQRHOGKVLV--LPASAPLAPGTIVLLTGGLAL 4742
 Qy 1291 -----STGDRDLDRILASAPARAALAEAFRTQVSVYLRTPEIKVG 1333
 Db 4743 GLHVARVLAQGVPHMVLTRRG--LDTPGAAKAVAEIEALGARITIAASDVADNNALEAV 4801
 Qy 1334 AEAL-----FTRLCMDSLMAVELNRRLEASLKLST 1365
 Db 4802 LQAIPTAEMPLQGVYHAAGALDDGVLDQTTDRFSFRVLAPKVTGAMNHLELTAGNDLAFV 4861
 Qy 1366 TELSTS-----PNTALLAQNLDAATASLERVAENLRAGVOND 1406
 Db 4862 LFSSMSGLSAGQSNTA--ANNTFLDLALAHRRRAEGLAQSILAWPMSD 4909
 RESULT 4
 AAB83974
 ID AAB83974 standard; Protein: 2153 AA.
 XX AAB83974;
 DT 06-AUG-2001 (first entry)
 DE Amino acid sequence of a type I polyketide synthase.
 XX Metabolic pathway operon; polyketide; polyketide antibiotic;
 KW type I polyketide synthase.
 XX Unidentified.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 1
 FT Misc-difference 1 /note= "Met encoded by GTG"
 XX
 EN W0200140497-A2.
 XX
 XX 07-JUN-2001.
 PD
 XX 27-NOV-2000; 2000WO-FR03311.
 PF
 XX 29-NOV-1999; 99FR-0015032.
 PR 07-JUN-2000; 2000US-0209800.
 XX
 XX (AVET) AVENTIS PHARMA SA.
 PA
 XX Jeanm P. Pernodet J. Guerinneau M. Simonec P. Courtois S.
 PI Cappellano C. Francou F. Raynal A. Ball M. Sezonov G. Tuphile K.
 FI Flostegard A.
 XX
 DR WPI; 2001-374849/39.

QY 1276 ---PEVO-----IRIPTHAGA-----1288
 DB 1801 RASPEVQQAIAVMDADVRTQADVQSDLELERYISSIDRARGVTHAAVLDALLNQT 1860
 QY 1289 -----GPESTD-----1294
 DB 1861 EAHPFNWAAKIDGAMNHLITRDCPLDHFVLFSSAGLGAPOGNYAANAFLDALAY 1920
 QY 1295 -----RDLD-----1299
 DB 1921 YRKAQGLPALSIGWAMSEVGLAAQDNRSRLALRGMENTLPOHGLALEQLNNSACH 1960
 QY 1300 -----RLASAPAPAAALEAF 1316
 DB 1981 VAAMPINVRQKQFYPKAAQSALPELHDDAASADAPNALRAHQSAEPQRTTLEEH 2040
 QY 1317 LRTQVSQVLTPEIKVGAELFTRIGMDSLMAVELRNREIASLKIKSTTFLSTSPNIAL 1376
 DB 2041 LQQQLARVLRIDSCITIDPLRPLKEKGFPSLMALEFRNRLTGLTLPATLIMGHPTLAG 2100
 QY 1377 LAQNLDLATLALSIERYAA 1396
 DB 2101 LAPHLASQWGLPLVEAQA 2120

RESULT 5
 ABG99871
 ID ABG99871 standard; Protein; 4038 AA.
 AC ABG99871;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE S. cinamomensis MonAIV/polyketide synthase multi-enzyme MONS4.
 XX
 KM Monensin; gene cluster; polyketide synthase; antibiotic;
 KM antihelminthic; insecticide; immunosuppressant; antifungal;
 KM antibacterial; polyether; mon BI; mon BII; mon CI; mon CII; mon H;
 KM mon RI; mon RII; mon T; mon AIX; mon AX.
 XX
 OS Streptomyces cinamomensis.
 PN WO200168867-A1.
 PD 20-SEP-2001.
 XX
 PF 30-MAY-2000; 2000WO-GB02072.
 XX
 PR 28-MAY-1999; 99GB-0012563.
 XX
 PA (BIOT-) BIOTICA TECHNOLOGY LTD.
 PI Leadlay PF, Staunton J, O'lynyk M;
 XX
 DR WPI; 2001-611393/70.
 DR N-PSDB; ABX04971.
 XX
 FT New DNA sequence encoding polyketide synthase, useful for the
 PT production of polyketides such as antibiotic monensin -
 XX
 PS Claim 6; Page 82-85; 212pp; English.
 XX
 CC The invention relates to a DNA sequence which is a fully defined sequence
 CC of 103521 base pairs appearing as ABX04971, or its variant, that it is
 CC not a sequence encoding all or part amino acids 1-920 encoded by mon AI
 CC as given in the specification. The DNA is the S. cinamomensis
 CC polyketide antibiotic monensin biosynthetic gene cluster. Also included
 CC are a recombinant cloning or expression vector comprising the gene
 CC cluster, a transformed host cell which has been transformed to
 CC contain the gene cluster (and is capable of expressing a corresponding
 CC polypeptide), a hybridization probe derived from the gene cluster
 CC (for identification and isolation of the same or analogous gene cluster,

CC e.g. one which binds specifically to a region of the monensin gene
 CC cluster selected from mon BI, mon BII, mon CI, mon CII, mon H, Mon RI,
 CC mon RII, mon T, mon AIX and mon AX), the use of the mon RI gene or
 CC variant and a monensin promoter to control expression of a heterologous
 CC gene in Streptomyces cinamomensis, a polypeptide encoded by a portion of
 CC the monensin gene cluster (preferably comprising mon BI, mon BII,
 CC mon AIX or mon AX or their mutants, alleles or variants), an epoxide
 CC enzyme encoded by mon CI, a cyclase enzyme encoded by mon CII, producing
 CC S. cinamomensis capable of enhanced levels of production of monensin
 CC comprising engineering it to overexpress the mon RI gene, S.
 CC cinamomensis containing multiple copies of the mon RI gene and/or its
 CC variants, expressing a gene heterologous to S. cinamomensis comprising
 CC transforming S. cinamomensis with DNA encoding a heterologous gene and
 CC expressing the gene under control of the activator gene mon RI or
 CC actII/orf4 and 13-proxyl erythromycin A. The processes and materials
 CC (enzyme systems, nucleic acids and vectors) are useful for preparing
 CC polyketides by recombinant synthesis. The polyketides are useful as
 CC insecticides, antibiotics, antihelmintics, antifungals, antibacterials or
 CC other pharmaceuticals. In particular the gene is useful for the
 CC production of monensin, an antibiotic polyether polyketide. The
 CC present sequence represents a protein encoded by the monensin gene
 CC cluster.

XX Sequence 4038 AA;

Query Match 32.8%; Score 2363; DB 22; Length 4038;
 Best Local Similarity 33.6%; Pred. No. 8.8e-179;
 Matches 647; Conservative 193; Mismatches 515; Indels 572; Gaps 47;

QY 2 ADRPIERABDPDAIVAGASCRLPGGVTDLSGEFTLLEGRDTCGRVPAER-WDAANFDP 60
 DB 1823 AGADVADTDPPAIVAMTCRFPGGVASPDLDMLDLARCKAMGAFPTDRMDLERLHP 1882.
 QY 61 DPDAKPTVTRASFSLDVACFDASFGISPREALRMDPARHLLVECMALENAAIAPS 120
 DB 1883 DDPHGTSTYDQGGFLPDAGDFDAFPFINREALAMPQOQLLEMSWEYLERAGIDPT 1942
 QY 121 ALVETGVTFIGIGPSEYEALPQATASAEIDAHGIGTSPVSGARISVALGRPCVA 180
 DB 1943 TLKGTPTGTVMYMHYAKSF--TDAQLEGVSYLATSQSMVSGRAVYLTGLEGPVAVT 2000
 QY 181 VDIRYSSSLVAVHLACSLNSGECSTLAVGVSIMLSPSTLWLSKTRALARDGCAFS 240
 DB 2001 VDTACSSSLVSIHLATQALRRGECDLALAGGVWADPDMFAGRSRGRGSLPDRCKAYA 2060
 QY 241 AADGPRGECCAVYVKRLSGARADGRIIAVIRGSAIINHDGASGLTVPNGSOEIVL 300
 DB 2061 AAADGVSEGVGLLERLSDARRRGRVYGVGRSAVNDGASNGLTAFNGPSQERVI 2120
 QY 301 KRALADAGCAASVGYEAHGTGTTLDPIEIQALNAVYGLGRDVATPLIGSVKYNIGH 360
 DB 2121 KQALASGGLSVDDVYVEGHGTGTTLDPIEIQALNAVYGLGRREDRPLWIGSVKYNIGH 2180
 QY 361 PEVASGITGLKLVLSLQHQIIPALHQAQALNPISW--GDLRLTVTRATPMD-WNTP 417
 DB 2181 TQAAAGVAGVTKVMAMRHGVVPASLHVDPSPHVEWDSGAVRLAV--ESVPAQVEGRP 2238
 QY 418 RRAVSSFGMSGTAHVYLEAP-----AATCTPPAPERAP-----LVLSARTASALD 467
 DB 2239 RRAVSSFGMSGTAHVYLEAP-----AATCTPPAPERAP-----LVLSARTASALD 467
 QY 468 AQAARLDHLETYSQC-IGDVAFSLATTSAMEHRLAUAATSREGRAALDAACQOTS 526
 DB 2239 DQALRLDFASDAFRAPLADVGSLLKTRALHHRVAVVGAEEAELALAEALATGEPH 2358
 QY 527 PGAVRSIADSSR---GKLAFLFTQGAQTLGAGGLYDVMSAFREAFDLCVRLFNQELDR 583
 DB 2359 AALVGPAQSGARVGGDDVVMFLFSGQSGLVGMAGGLYERFPVFAAFEDVEVGL----LEG 2414
 QY 584 P-----LRVMAAEPAASVDALLDOTAFQPLLFTEVYALALWMSWGPPEPLVAGHS 636
 DB 2415 PLVGAGGLREVVRGPRE-----RLDHTVMAQGLFALQVGLARLWESVGVPRPDVLAGHS 2470

```

QY 637 IGEIIVAAACVAGVFESEDATVFLVAAARGRLMQLPAGAMVSTIEAPDAVAAVAPHAASV 696
DB 2471 IGEIAAAHVAGFDLADACRVGAPRRMLGSLPEGACAKCAVQATPELADV--DQSAVS 2528
QY 697 IAAVAPQOVNIVAGAGQPVHAIIAAAMAAAGARTKALHSHAFSPMLAPMBAFGFVAES 756
DB 2529 VAAVNTPBSTVLSGFSDVDYRAGVWRERGRKTKLSVSHAFSHALMELMELFEALRG 2588
QY 757 VSYRPSIVLVNLSGKACTDEVSSFGVWRHAREVVRPADGVKALHAAAGCTFVEVGPK 816
DB 2589 VAFRPSIPIIMSNVSGERAGEEITDPEYWRHVRNVAFLFQPAI-AQVADSAGVFVELGPA 2647
QY 817 STLGLVPRCM--PARAPALLASRRAG--RDEPATVLEALGVMAGLVSMAGLSPSSGR 873
DB 2648 PVLTTAAQHTLDBSDSQSVLSLAGEPESAFVEAMARLHTGAVADMVSLF--AGD 2705
QY 874 RVP-----LPTYPWORERYWIDTK----- 892
DB 2706 RVPGLVELPTVAFQREHFWLSRSGGDAATLGLVAAQHPLLGAAVEFADRGCLLTGRL 2765
QY 893 -----ADDAARDDRRAFGA-----GHDE----- 910
DB 2766 SRSGVSWLADHVAGAVLVPAGALVEMALRAGDEVGCTVEBELMLQAPLVEASGLRVQ 2825
QY 911 --VEEGG-----AVRGDR-----RSARLD-----HSP----- 931
DB 2826 VVEEAGDEDRRGVQIYSRPDADAVGGDDSWICHATGVLSPEPSARLDLTGGVWPAGAE 2885
QY 932 -----PES----- 934
DB 2886 PLVDGFFVAQAGEAGVGPAPRGLRAVWRHGODLLAEVVLPEAGAHNDGYGIHPALLDA 2945
QY 935 ----- 934
DB 2946 TLHPLLAARFMDSEDDQLYVFGWAGVSLRAVGATTVRRLRPVGEVSVDGLSVTVTDA 3005
QY 935 -----GR 936
DB 3006 TGGPVLVSUDSLQTRPVKPSQLAAQOPDVRCGLFTVEMTPLPOTDADGEADMVVLSDGVR 3065
QY 937 -REKTEAAG----- 944
DB 3066 LADVVAAGGEAPMAVVAVPDASVGDGRGRLDRLVERVLSLVQEFLLPELASRLIV 3125
QY 945 -----DRPRL----- 950
DB 3126 VTRGAVATGVDGDGDVDAASAAWGLVRSAGSENPGRFILLDDVDDGDGQDFDLGRHLR 3185
QY 951 -----EIDEP-----GVLDHLVLRTER--APG----- 972
DB 3186 HATLHAAEELDEPQLALREGTL--VPRLTQARQSAELVPRGEPARMLRWHDGSLDA 3243
QY 973 -----LGEVEIAVDAAGLSFNDVQLALGMVDDLPCKRNPPLLIGECACR 3295
DB 3244 LAAVACPEALBPLAQGVRIAAHAGINFRDVLVNLGMVP-----AYGAMGEGAGV 3295
QY 1019 IYAVEGCVNGLVYGCVPVIALSAGAFTHYTTSAALVLRPPQALSAIEAAMVAYITAMY 1078
DB 3296 VTEVSPREVTHVSVGKRVKGVFEGARQPVVIAARWTPVPQGMWRKAAAGIPAAFLTAMY 3355
QY 1079 ALDRITARLOPGERVLIHAATGVGLAAVQMAOHVGAENATGTPEKRAYLSLGV--RY 1136
DB 3356 GLVELLGLAGGERVLIHAATGVGMAVQIAHVGAEVPAITA-SPEKHAIVLEMGIDAAH 3414
QY 1137 VSDSDFRVDADVRAWTGEGGVVNLNSGELIDKSFNMLRSHGRFVELAGKRDYADNO 1196
DB 3415 RASSDILAEFGTFRATGTGRGMDVNLNSAGEFIDLSLFLDDGGRFLMGKTDVRAABE 3474
QY 1197 LGLRPLRLNLSLVLDLRGMMLERPARVALLBELGLIAAGVTPPPIATLPIARVADA 1256
DB 3475 VAAEH--ADVSTAYDLVD--AGPDRISNMLDKVELLEASERLKLPLVRSWPLDKACEA 3530
QY 1257 FRSMQAQHLGLVLTIG--DPEVQIRLPTHAGAGPST-----GDRDL--DRLA 1302

```

```

DB 3531 FRFMSQAHTGKLVLEIRPALDPECTVIVTGTGALGVVAEHLVRENGVHLLASRRG 3590
QY 1303 SAAPARAALAEALRTQVSGVLTPEIKVGAELFTRLGM-DSLMAVELRNRIEASLKL 1361
DB 3591 PEAPG-----SDELASKLTGLGAEVTIVAADVSDPASVVELVGTDPSPHPL 3636
QY 1362 KLSTFELSTSPNIALAQNLDALATALS---LERV-----PAENTL-----RAGVQN 1405
DB 3637 -----TGVAHAAGVLEDGVVTAQTPGLARVMAKAAANLHEATRENKLGFLV 3686
QY 1406 DFVSSGA 1412
DB 3687 VFSSAAA 3693

RESULT 6
AAE10145
ID AAE10145 standard; Protein; 5435 AA.
XX
AC AAE10145;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin gene, NysJ.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX antifungal; antibiotic.
XX
OS Streptomyces noursei.
XX
PN W0200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UNO-) UNIV NORGES TEKNIISK NATURVITENSKAPELIGE.
PA (SNTE) SINTEF STIPELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FUAE/) FUAERVIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR MPI: 2001-557614/62.
DR N-PSDB: AAD17186.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
XX useful as antibiotics and antifungals -
XX
PS Claim 15; Page 260-263; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
XX cluster encoding a modular type I polyketide synthase (PKS) enzyme
XX involved in the biosynthesis of the macrolide antibiotic nystatin.
XX The nystatin PKS is useful as antifungal antibiotics. The present
XX sequence is a PKS encoding Streptomyces noursei nystatin gene, NysJ.
XX
SQ Sequence 5435 AA;
XX
Query Match 32.7%; Score 2355.5; DB 22; Length 5435;
Best Local Similarity 34.2%; Pred. No. 5.6e-176;

```



```

Db      3092 VLSIDVGSIDVSVSAAGSEAPWAVAPVAGSAGGIGAGDRREGIDGRLVVERVLSVOE 3151
Qy      923 -----RSARLDHP-----930
Db      3152 FLABELAESRLVLTIRGAVATGDDGDVDASAAAWGLVKSQSENPGRIILLDVMDD 3211
Qy      931 -----PPESGR-----937
Db      3212 VDVIDVMDVDVDVDVDVDGNGSDLDLDPLNGRLPHATIRHAAELEDQQLARDDQ 3271
Qy      938 -----EKVEBAG-----DRPFLE-----IDFQVLDHLVLRVTERARGL 973
Db      3272 LLVPLRVATGGGLVVAFTDRAMRLDKSAETLESVAAPVAPGVMEPL-----GP 3321
Qy      974 GEVEIAVDAAGLSFDVQOLAGMVPDDLPGKNPPLLGBEAGRIIVAGEGVNLVVGQ 1033
Db      3322 GQVRIGIHAGINFPDVLISLGMVPGVG-----LGBGAGVVTETGPDVTHLSVD 3373
Qy      1034 PVIALSAGAFATHTVTSALVLPRLPQALSAIEBAAMPVAYLTAWYALDRIALQGERVL 1093
Db      3374 RVMGVLHGSFQPTAVADTRMVAPEQGMVMQAAAMPAYILTAWYGLVELAGLKAGERVL 3433
Qy      1094 IHAATGVGLAAVQAVQAVGAEVHATAGTPEKRAYLESIGV--RYVSDSRSDRPVADVRA 1151
Db      3434 IHAATGVGMAAVQIARHLAGAEVFATA--SAAKHVLEEMGIDAARASRDIAFEDTRQ 3492
Qy      1152 WTGGGVVVVNSLSGELLIDKSFNLRSKRPVELGKDCVANDQLGLR--PFLRLSLSTL 1210
Db      3493 AIDGSGMVVNSLISLGEFIDASLRLLGDGRLGKTDVTPTEVVAEYPGVITYVDL 3552
Qy      1211 VDLRGMLEPRPARVRLLEELGLIAAGVTPPTPLPIAVADAFRSMAQOHIGKLV 1270
Db      3553 VTDAG-----PRIVMSELGERFASGALDLPVRSWPLDRAARAFRMSQAKTGKLV 3607
Qy      1271 LTLG---DPEVOIRIPTHAGA-----GPSTGDDLLDR 1300
Db      3608 LDVPAPLPDGTVLITGGTALGQVVAEHLVREMGVRLHLLASRGLDAPGSG--ELADR 3665
Qy      1301 LASAPARAAALE 1314
Db      3666 LSDLAGAVTVAAAD 3679

```

```

FT      /label= ACP3_domain
FT      /note= "Acyl carrier protein (ACP) domain"
FT      1757..2180
FT      /label= KS4_domain
FT      /note= "Ketosynthase (KS) domain"
FT      2291..2603
FT      /label= AT4_domain
FT      /note= "Acyltransferase (AT) domain"
FT      2617..2818
FT      /label= DH4_domain
FT      /note= "Dehydratase (DH) domain"
FT      3124..3371
FT      /label= KR4_domain
FT      /note= "Ketoreductase (KR) domain"
FT      3407..3480
FT      /label= ACP4_domain
FT      /note= "Acyl carrier protein (ACP) domain"
FT      3501..3924
FT      /label= KS5_domain
FT      /note= "Ketosynthase (KS) domain"
FT      4032..4346
FT      /label= AT5_domain
FT      /note= "Acyltransferase (AT) domain"
FT      4360..4561
FT      /label= DH5_domain
FT      /note= "Dehydratase (DH) domain"
FT      4953..5239
FT      /label= ER5_domain
FT      /note= "Enoylreductase (ER) domain"
FT      5248..5495
FT      /label= KR5_domain
FT      /note= "Ketoreductase (KR) domain"
FT      5528..5601
FT      /label= ACP5_domain
FT      /note= "Acyl carrier protein (ACP) domain"
FT      5623..6046
FT      /label= KS6_domain
FT      /note= "Ketosynthase (KS) domain"
FT      6165..6478
FT      /label= AT6_domain
FT      /note= "Acyltransferase (AT) domain"
FT      6492..6704
FT      /label= DH6_domain
FT      /note= "Dehydratase (DH) domain"
FT      7038..7281
FT      /label= KR6_domain
FT      /note= "Ketoreductase (KR) domain"
FT      7315..7388
FT      /label= ACP6_domain
FT      /note= "Acyl carrier protein (ACP) domain"
FT      7408..7831
FT      /label= KS7_domain
FT      /note= "Ketosynthase (KS) domain"
FT      7939..8253
FT      /label= AT7_domain
FT      /note= "Acyltransferase (AT) domain"
FT      8267..8470
FT      /label= DH7_domain
FT      /note= "Dehydratase (DH) domain"
FT      8812..9086
FT      /label= KR7_domain
FT      /note= "Ketoreductase (KR) domain"
FT      9120..9193
FT      /label= ACP7_domain
FT      /note= "Acyl carrier protein (ACP) domain"
FT      9214..9637
FT      /label= KS8_domain
FT      /note= "Ketosynthase (KS) domain"
FT      9758..10072
FT      /label= AT8_domain
FT      /note= "Acyltransferase (AT) domain"
FT      10086..10289
FT      /label= DH8_domain

```

```

RESULT 8
AAE10129
ID AAE10129 standard; Protein; 11096 AA.
AC AAE10129;
XX
XX
XX 29-NOV-2001 (first entry)
DE Streptomyces noursei nystatin gene, NysC.
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW antifungal; antibiotic; PKS type I.
XX
OS Streptomyces noursei.
XX
XX Key
XX FH Location/Qualifiers
XX FT 35..455
XX FT /label= KS3_domain
XX FT /note= "Ketosynthase (KS) domain"
XX FT 546..858
XX FT /label= AT3_domain
XX FT /note= "Acyltransferase (AT) domain"
XX FT 872..1073
XX FT /label= DH3_domain
XX FT /note= "Dehydratase (DH) domain"
XX FT 1381..1628
XX FT /label= KR3_domain
XX FT /note= "Ketoreductase (KR) domain"
XX FT 1662..1735
XX FT Domain

```

FT /note="Dehydratase (DH) domain"
 FT 10657..10904
 FT /label=KR8_domain
 FT /note="Ketoreductase (KR) domain"
 FT 10939..11012
 FT /label=ACP8_domain
 FT /note="Acy1 carrier protein (ACP) domain"
 XX WO200159126-A2.
 XX 16-AUG-2001.
 XX 08-FEB-2001; 2001WO-GB0509.
 XX 08-FEB-2000; 2000GB-0002840.
 XX 10-APR-2000; 2000GB-0008786.
 XX 14-APR-2000; 2000GB-0009387.
 PA (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
 PA (SNTE) SINTEF STIPELSEN IND TEK FORSK.
 PA (ALPH-) ALPHARMA AS.
 PA (SINV-) SINEVENT AS.
 PA (DZIE/) DZIEGEMSKA H.
 PA (ZOTC/) ZOTCHEV S B.
 PA (SEKU/) SEKUROVA O N.
 PA (FJAE/) FJAEVRIK E.
 PA (BRAU/) BRAUTASET T.
 PA (STRO/) STROM A R.
 PI Zolchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
 PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
 XX MPI; 2001-557614/62.
 DR N-PSDB; AAD17184.
 XX New nystatin polyketide synthase polynucleotides and polypeptides,
 PT useful as antibiotics and antifungals -
 XX Claim 15; Page 170-176; 266pp; English.
 CC The present invention relates to the cloning and sequencing of the gene
 CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
 CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
 CC The nystatin PKS is useful as antifungal antibiotics. The present
 CC sequence is a PKS type I encoding Streptomyces noursei nystatin
 CC gene, NysC.
 XX
 SQ Sequence 11096 AA;
 Query Match 32.3%; Score 2331.5; DB 22; Length 11096;
 Best Local Similarity 32.9%; Pred. No. 1,4e-175;
 Matches 629; Conservative 162; Mismatches 519; Indels 583; Gaps 34;
 QY 10 AEDPIAIVGASGRLPGGVLDLSGFWTLBGSRTVGRVPAER-WDAAMFDDPDPAAGKT 68
 DB 3499 ADDPIYVGMACRFPGGVTSPEBMLRVLDGDITTFPTNRRMDLNDLYDPREHFGTS 3558
 QY 69 PVTTRASFLSDVACFDASFFGDISPREALRMDPAHRLLEVCWEALENAALAPSAIVGTETG 128
 DB 3559 YTRSGGFLHEAGFEFDPAFGMSPREALATDSOQRLTLESSWEAIEBAGIDPLTRGSATG 3618
 QY 129 VPIGIGPSEYEALPATASAEIDAAGGIGTMSVAGRIISYALGARGCVAVDTAYSSS 188
 DB 3619 VFGVWYISDYGLT---GGEFEGFGQGSAGSVASALGEGPAVTVDTACSSS 3674
 QY 189 LVAVHLACOSLRSGECSTALAGVSLMLSPSTLVMLSKTRALARDRCFAFSEADGFG 248
 DB 3675 LVAAHMAAALRAGECSLLAGVTVMTSTPSTVEBSRQGLAPDRSRSAFAAALAGVM 3734
 QY 249 GEGCAVVLKRLSGAPADGRIIATVIRGSALNHGASGGLTPNGSSQEIYVKRALADAG 308
 DB 3735 SEGVGILVIERQSDAVRNGHEIIATVIRGSALNHGASGGLTPNPGSQOQVRVIRQALASG 3794

QY 309 CAASGVYEAGGTGTTGDDPIEIOALNAVYGLGRDVATPLLIGSVKTNLGHPEYASGTT 368
 DB 3795 ISTDADVAEAGTITTDPIEAOALATIGRDPDENPMLISLKNIGITQAAAGVA 3854
 QY 369 GLKLVLSLQHQIIPAHHAQALNPRISMGDLRLTVTRARTWPDMPNTPRAAGVSSFGMS 428
 DB 3855 GVIKVMAMRHGVLHQTLLHVDAPSHTVMSVGAVELLIEQIAMPETGVRRAAGVSSFGIS 3914
 QY 429 GTNAHVLE-----EAPATCTPPAPERRPAEL-LVLSARTASALDAQARLDHETPS 482
 DB 3915 GTNAHVIVEQPALVSSPAA---EPGRBPGVVPLPLSKSPALDQAARLLAGLAERPA 3971
 QY 483 QCLGVASFATTRSGAMEHRLAVALTSREGRLAALDAQAQGTSPGARSIADSSRGKLA 542
 DB 3972 LRPDLGSLATTRSAFPHRAEVLTDAADVALRTALLAAADADISAV--VQDTTGCHA 4029
 QY 543 PLFTQGAQTLMGRGLYDVMSAFREAPDLCYRLFNOEL--DRPLREYMAEPASVDAAL 600
 DB 4030 VLFSGQSGORLGMSELVERFPVFAELDVAIIDHLDALPAQASLREVMWGQ---DVEL 4085
 QY 601 LDQTAFTQPALFTFEYALAAALMRSGVEBELVAGHSIGELVAACVAFSLEDVAFVVA 660
 DB 4086 LDETGMTQPALFAVVALFRLVESGWVRPDPVAGHSIGEIAAAHVAGFSLDACLRLVAA 4145
 QY 661 RGRLMQALPAGGAWSIENPREADVAAAVAPHAASVISAANVAPDOVVLAGAQPVHAIAA 720
 DB 4146 RATLMQALPTGAMATAIOADEVTQHLTD--DIVSIANVGTFTVVSAGAESARTYAD 4202
 QY 721 AMAARGARTKALHVSAPHSPMAEMLEAFGRVAESVSYRRPSIVLSNLSGAKCT-DEV 779
 DB 4203 RLAEGRKTRRLRVSHAFHSPMDMLEAFRAVAGLSYATPTLTPVSNLTGRLTADDL 4262
 QY 780 SSPGYWVHARVAVFADGVYKALHAAGCTPVEVPKSTLGLVACMP-PARPALASS 838
 DB 4263 CSAEYMAHVRBAVFAFGVSTLENEGVTTLFELPDDVLNMAQOSLTGD--ATVPAL 4320
 QY 839 RAGRDEPATVLEALGLMAVGLVSMAGLFP--SGRRVPLPTYPMQRRRYW----IDTK- 892
 DB 4321 RKRDEETSALTALHLHTAGLRVDMMAFFPAGSGATRVLDLPYTAQNHATYPTGLPLPAH 4380
 QY 893 -----ADDAARGRRAPG----- 905
 DB 4381 AAAYGLTAAEHPLNGSVELAEGBGLFTGRLSQSHPMLDHVMAGVLLPGLALLELA 4440
 QY 906 -----AGHDEV-----EGAV----- 917
 DB 4441 FRAGDEACQDRVEELTLAALVLPFPGAVQTVQVRGVADTGRRTVTVHSRBEHATDVSW 4500
 QY 918 -----R 918
 DB 4501 TQHATGTLTMSGAPADTGFDATAMPRADEPLATDCCARFTTLGFAVGPVQGLRAAMR 4560
 QY 919 GGD----- 921
 DB 4561 AGDVLYAEVLAESTGDEBATAFGILHPLALDAAHSLVAHEGEESNGCLPESWEGATLYA 4620
 QY 922 ----- 921
 DB 4621 TGATALRVRLPFTGDSGVAAIAVADTAGRPVAAIDNLVSRVSGDQLTGAAGLARDALF 4680
 QY 922 ----- 921
 DB 4681 TLDMNPVENLVENPVENBTGGHAQODGRPAALATVALVAGADDTALAADTLTAAGIHIT 4740
 QY 922 ----- 921
 DB 4741 LHPDLTTLATDADVKTVLPLTGTGTGTGTESTDGTGTAESDASAPSPAEVAHT 4800
 QY 922 -----RSGAR----- 926
 DB 4801 ISTAALALVQEWTAQERFAGSRFLAVTTGTATAGGTDVMDVAAAVWGLVNSAQSEAPDT 4860
 QY 927 ---LDHPPESGRREKVENA----- 943

```

Db      4861  FVLIRDPGPAGTHRTAAERGQLLRALHTDEPOLALRDGVLAAARLAREDTAAALTP 4920
      944  -GDRPRLIEDPGLDHLVLR-VIERRAPGIG-EVEIIVDAAGSLFNDVOLALMVPDD 1000
      4921  PADRRMRRLSTAKSGINGLALTPYPALALPLTGHEVRVVRAGLNFRLVALLMVPDD 4980
      1001  LFGKPNPPLLLGECGARIIVAGGVGLVGPVIALSAGAFATHTVTLSALVLRPPOA 1060
      4981  DVGSG-----FGSEAAGVVEVGPETVGLAPDQVMGMTSGSGSLAVDADARLARLPED 5034
      1061  ISAIEAAMPVAYLTIAVALDRIARLOPGERVLIHAATGCVGLAAVQAQHVGAEVHATA 1120
      5035  WEMETGASVPLVFLFYALKEGLRAGEKYLVAHAGAGVGAALQIARHGAIEVFATA 5094
      1121  GPPEKRAYLESIGVR--YVSDRSDFVDVRAAMGCEVDVNLSTGSELDKSFNLIR 1178
      5095  -SEGRMDVRSLSGVADDDHTASSRTLDFFAFAEAVAGDRGLDVNLNLAGDFVDAISRLLG 5153
      1179  SHGRFVELGKRDYADNOLGLRPLRLNLSFSLVDLRGMLEPARVARALLLELLGLIAG 1238
      5154  DGRFLEKTKDIRADSVF-----DGLSYQSFDLAMVY---PETIGTIALMLDFRFG 5205
      1239  VETPPPIATLPIARVADAFRMAQAQHLGKLVTL---GDPEVQIRIPTHAGAG----- 1289
      5206  ALRPLFVRTWVRAHAKDAFRFMSMAKHIGKIVILPRSKPBGTVLVY--GGTGGIGLV 5263
      1290  -----PSTGDRDLRLRLASAPARAALEFLRTQVSOVLATPELKVAELFTRLGM- 1343
      5264  ARLHVRSCGVRHLILTSRSGVGAAGAGLVAELES-----IGARVVVAACDVG 5311
      1344  DSLMAVELNRLEASLKLSTTFLSTSPNIALLAQNLLDALATALSERVA 1396
      5312  DGSAAVELVAGVSESTPLSA-----VHAAGVLDGVVGLSPERLAA 5354

```

RESULT 9
 ABP57679
 ID ABP57679 standard; Protein; 2149 AA.
 XX ABP57679;
 DT 22-JAN-2003 (first entry)
 DE Saccharopolyspora busb butenyl- γ -spinosyn biosynthetic gene product.
 KW Butenyl γ biosynthetic enzyme; PKS; polyketide synthetase; macroliide;
 KM metabolite; spinosyn.
 OS Saccharopolyspora sp..
 XX
 PN WO200279477-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-US09968.
 XX
 PR 30-MAR-2001; 2001US-280175P.
 XX
 PA (DOWC) DOW AGROSCIENCES LLC.
 PI Hahn DR, Jackson JD, Bullard BS, Gustafson GD, Waldron C;
 PT Mitchell JC;
 DR WPI; 2003-058434/05.
 DR N-PSDB; ABV75557.
 XX
 PT New butenyl-spinosyn biosynthetic genes, useful for increasing the
 PT production of butenyl-spinosyn insecticidal macroliides, or for changing
 PT the metabolites or products produced by spinosyn-producing
 PT microorganisms -
 PS Claim 1; Page 134-141; 218pp; English.

```

XX      The invention relates to a novel DNA molecule comprising a DNA sequence
CC      that encodes a butenyl-spinosyn biosynthetic enzyme, a butenyl-spinosyn
CC      PKS (polyketide synthetase) domain, or a spinosyn PKS module. The
CC      butenyl-spinosyn biosynthetic genes are useful for increasing the
CC      production of butenyl-spinosyn insecticidal macroliides. The genes are
CC      also useful for changing the metabolites or products produced by
CC      spinosyn-producing microorganisms. The present sequence represents one of
CC      the butenyl-spinosyn biosynthetic polypeptides of the invention.
SQ      Sequence 2149 AA;
Query Match 32.2%; Score 219.5; DB 24; Length 2149;
Beet Local Similarity 32.8%; Pred. No. 1e-175;
Matches 610; Conservative 198; Mismatches 547; Indels 507; Gaps 32;

```

9 AAE--DPAIYAGSRLPGVYIDSGFTLLGSRDITGRPAER-WDAAMFDDDPAP 65
 30 AAEKGDPIAIVAMSCRYPGVQVSSPEDLWQLAAGVDAISEVPGDGMOLAGFPDSDRP 89
 66 GKTPTTRASFISDVACFDASFFGGISPREALAMPARHLLFECWEALENAAIAPGALVGT 125
 90 GTSTACAGGFLQGVSEFDAGFFGISPREALAMPQRLLEVAWEVFERAGLEGSTGGS 149
 126 ETGVFTIGIPSEYEALPQATASAEIDAAGIGTWPVSGAGRISYALGLRGPVAVDTAY 185
 150 RVGVFTVINGDQVASMW--RTPPEVAGHVLTGGAAALISGRVAVSFGFEGPAVTDTAC 207
 186 SSSLVAHLAQSLRSRGECSTALAGVSLMSPSTLVLSKTRALARGRCKASAEADG 245
 208 SSSLVAHLAQSLRAGCDLALAGVYVMTSTPAFTEFSRORGLADGRCKSFMAADG 267
 246 FGRGCAVNVVLKLSGARADSDRIAYIRGSATINHOGASSGLTPNCSOEIYKRALA 305
 268 TGMGEGAGLLLERISDARRNGHRVLAVRGSAVNOGASGLTAPNCSQARVITQALA 327
 306 DAGCAASSVGYVEAHGTGTTGDDPEIIOALNAVYGLGRDVATPLLIGSKTNLGHPEYAS 365
 328 SAGLSVSDVDAVEAHGITTGRDPEIAQALATATGRDDBPARPLMGSVKNSIGHTQAAA 387
 366 GITGLKTVLSLQHQIPAHLLHAQALNPRISWGDLRLTVTRARTPWPDMTPRRAGVSF 425
 388 GVAGVTKVMAMRRHQLPRTLHVDAFSPEDVMSAGTVQLITEMMLPESGVRRAGVSSF 447
 426 GMSGNAHVLEEARAACTPPAPER-----PAELVLSARTASALDAQARLDHET 479
 448 GISGTNAHVILBQPTGETRQAGPDGSGVDVPPVPMVSGKTPDALGADDTLMSYDD 507
 480 YPSQCLGDAFSLATTRSAMEHRLAVATSRREGRLAALDAAACQGTSPGAVRSIADSSRG 539
 508 RVDVPSLDIAVSLMTRFALDBRAVVLGPDRETLLSGKALSAGHEAGVVTGSGTG-G 566
 540 KLAFLFTGQAGVLTGMGRGLYDVWSAFREAFD-LCVRL--FNOELDRPLREVMKAEPAS 595
 567 RIGFVFSGGQGMIGMGRGLYRAFPVFAADEFCAEIEAHAGGEVG--VRDVVFGS--- 621
 596 VDAALDQTAFTQPLFTFEYALALAMRSKVEPELVGHSIGELVAVCAVGVSPLEAD 655
 622 -DAQLLNFTLMAQSGFLAQVGLKLKLDLSWGRPSAVVGHVSGELAAAPAGVSLSDA 680
 656 FLVARGLMQALPAGAMVSIAPADVAVAVPAHASVSIANVANAPDQVVIAGOPV 715
 681 RLIVAGRALLMQLPBGCGMLAVAGEBQRLPLADHGRVGLGANVNAESVYLSQDRVL 740
 716 HAIAAAMARQARTKALHVSHPHSLPMAFMLAEAFGRVAESVSYRPSIVLVSLSGA- 774
 741 DDIAGRILDGQGVRTFRLVSHAFHSYRMDPMLDEFAEIAARADVRCBLPIVSTLYGLKD 800
 775 CTDEVSSGVYVRRHAREVRPADQKALHAAGAFVGVGKSTLLGLVPCM-----PDA 830
 801 DAGRMSGPDVWRQVREFVRPADGQALVEHDVAIVTIGPDGASALAIQECVAASDOSR 860
 831 RPALASSRGRDDEPATVLEALGILMAVGLVSNAGLFP-SGGRVRPLPTYPMQRRERYMI 889

```

Db      861 RYAAVPMNRNDEAONLTALLAQVHVRGAVDMRSEFFACTGAKOVELPTVAFQRQRYL 920
QY      890 DTKADDAAAGDRRA-----PGAGHDEVEEGG----- 915
Db      921 EESDSGDVVGAGLAGAEHPLGAVVPAAGDEVLTGRISVGTHPWLAEHRVLGEVIVPG 980
QY      916 ----AVGGDR-----RSARLDHP-----PESGRREV----- 940
Db      981 TALLELTALHAGERLGCERVEBELTEAPLVLPKRGAMOVOLRVGAPENSGRRPWLVS RPE 1040
QY      941 ----EAGD----- 945
Db      1041 GAADHDWTRHATGRLAPGGGAAGDLADMPAGALPVDDEFRDLAEHGLEVPYIFQGL 1100
QY      946 ---- 945
Db      1101 KAMRQGEVYAALPGTDSGFGVHPALLDAAALHATAVRMDMGLPFGWEGVCIHAR 1160
QY      946 ----RP----- 947
Db      1161 AASALRVVPAAGDAKSLVCDGTGRPVISVDRLVRSAAAGRTGARROHRAHRLYRLG 1220
QY      948 ---- 947
Db      1221 WPTVOLPTSAQPPSCVLGTSEVSDMOVYPDLRLTALLDAGAEPPGVIAPTPGGQ 1280
QY      948 ---- 947
Db      1281 TADVESTREHALLDVQGLADRLNDRLFLVTRGAVAVEPGEPTDLAQAALMGLLRST 1340
QY      948 ---- 947
Db      1341 QTEHPRFVLVDVAEPAQLPALPGVLAGEPOLAIRGGAHAPRIAGLGGDDVLFPDS 1400
QY      948 --FRLEIDEPGVLDHLVLRTERRAP--GLGEVEIADVDAAGSFNDVOLATGMVDDLP 1002
Db      1401 MGRMLEATSPGTLIDGLAL--LDEPAATASLGDGQVRITAMRAAGVNFADALIALGMVGG-- 1456
QY      1003 GKRPPELLIGBECAGRIVAVGEGVGLVVGQPVIALSAGAFATHTVTTSALVLRPQALS 1062
Db      1457 ----AASLGGEGAGVAVETGPGVTGLAPGDRVMGMI PKAFGLAVADHRMVTIRIPAGWS 1511
QY      1063 ATEAAMPVAVITAMVALDRIRIARLORGEVLIHAATGVGLAIVQACHVGAEVHATCT 1122
Db      1512 FQAASVPIVFLTAYTALVDLAGLRGESLUVHSAAGVGMAATQLARKHIGAEVATASE 1571
QY      1123 PEKRAYLESIGVRYVSDSRDRFVADVRAMTGGEGVDVVLNLSGELIDKSFNLLRSHGR 1182
Db      1572 DKMQAV--ELTRERLASSRTPCEKQFLGATGSGVDVVLNLSIAGDFADASLRMLPRGR 1629
QY      1183 FVELGKRDCTADNOLG-LRPLRLNLSFSLVDLRGMMLERPARYRALLBELGLIAAGVTT 1241
Db      1630 FLELGKTDVRDPEVEVADHAPGVSYQAFDVE-----AGFORGEMLDELVELFEESGVLE 1683
QY      1242 PPPIATLPFARVADAFRMAOAHGLKVLTLTGDPVQIRIPFHAGAGP-----STGDD 1296
Db      1684 PLDPVTAWDYRQAPERARHLISQARHVGKLVLP-----PMDWAGVTLVLTGGTG-- 1732
QY      1297 LLDRLASAPAAAPAAALAEFLRTQVSQVLRTEIKVGAE--ALFTRIAMD-SIMAVELELN 1353
Db      1733 ----ALGAEVAVARHLVTEHGVNRLVLSRRGPAASGAELVQLTLYGAEVSLQACDVAD 1787
QY      1354 RIEBASIKLSTFTSTSPRIALLAQNLLDALATALSERV-----AAENLRAGVQ 1404
Db      1788 R-ETLAKVLAGIPDEHTLTAVVAAGVLDGVAESITAOQLDHLVRPKVGDARNLIELTA 1846
QY      1405 ND 1406
Db      1847 PD 1848

```

```

AAV39298
ID AAV39298 strandcd; Protein; 2152 AA.
XX
AC AAV39298;
XX
DT 01-DEC-1999 (first entry)
XX
DE SpnB a polyketide synthase.
XX
KW Spinosyn biosynthetic enzyme; open reading frame; ORF; insecticidal
KW microtides; arachnid; nematode; insect; polyketide; polyketide synthase;
KW PKS; extender module; initiator module; acyl transferase domain; AT;
KW acyl carrier protein; ACP; beta-ketosynthase domain; KS; KR;
KW dehydratase domain; DH; enoyl reductase domain; ER; beta-ketoreductase;
KW insecticide.
XX
OS Saccharopolyspora spinosa.
XX
FH Key Location/Qualifiers
FT Domain 1..424
FT /label= KS2
FT /note= "Beta-ketosynthase domain: part of
FT /note= "extender module 2"
FT Domain
FT 536..886
FT /label= AT2
FT /note= "Acyl transferase domain: part of extender
FT /note= "module 2"
FT Domain
FT 892..1077
FT /label= ACP2
FT /note= "Acyl carrier protein domain: part of extender
FT /note= "module 2"
FT Domain
FT 1338..1683
FT /label= ER2
FT /note= "Enoyl reductase domain: part of extender
FT /note= "module 2"
FT Domain
FT 1687..1866
FT /label= KR2
FT /note= "Beta-ketoreductase domain: part of extender
FT /note= "module 2"
FT Domain
FT 1955..2034
FT /label= ACP2
FT /note= "Acyl carrier protein domain: part of extender
FT /note= "module 2"
XX
PN WO9946387-A1.
XX
PD 16-SEP-1999.
XX
PF 16-FEB-1999; 99WO-US03212.
XX
PR 09-MAR-1998; 98US-0036987.
XX
PA (IDWC ) DOW AGROSCIENCES LLC.
PI Balazs RH, Broughton MC, Crawford KP, Madduri K, Merlo DU;
PI Treddway PJ, Turner JR, Waldron C;
XX WPI, 1999-551414/46.
XX N-PSDB; AA221501.
XX
PS Claim 1; Page 96-103; 190pp; English.
XX
CC This is the amino acid sequence of the product of the spnB gene. The
CC protein is involved in spinosyn biosynthesis. The spnB gene is one of 23
CC genes and open reading frames contained in an 80kb DNA sequence
CC AA221501. Spinosyns are insecticidal microtides which are useful for the
CC control of arachnids, nematodes and insects. Biosynthesis of spinosyns
CC occurs via stepwise condensation and modification of carboxylic acid
CC precursors generating a linear polyketide which is modified further. The
CC DNA sequence contains a central region of approximately 55kb which has

```


homology to the DNA encoding the polyketide synthases (PKS) of known macrolide producers. The spinosyn PKS DNA region consists of 5 ORFs with stop codons at the end of acyl carrier protein (ACP) domains. Together the PKS polypeptides (AAV39297-Y39301) of which this sequence is one, form a complex consisting of an initiator module, spnA, and several extender modules spnB-spnE. Each extender module adds a specific acetyl-Co-A precursor to a growing polyketide chain, and modifies the beta-keto group in a specific manner. A module in a PKS polypeptide consists of several domains with specific functions. The initiator module has an acyl transferase (AT) domain, and an acyl carrier protein (ACP) domain. The extender modules have the same domains plus a beta-ketosynthase (KS) domain, and an enoyl reductase (ER) domain. The last extender module terminates with a thioester domain. The products of the genes present in the upstream region of the PKS genes have been assigned names spnF-spnS. AAV39302-Y39315 and are responsible for different modifications in spinosyn biosynthesis. There are also two ORFs ORF15 and ORF16 present immediately upstream of spnS, producing polypeptides AAV39316-Y39317, and two ORFs ORF1 and ORF2 present downstream of the PKS region producing polypeptides AAV39318-Y39319. The genes are useful to improve yields of spinosyns, and for creating new spinosyns e.g. by mutagenesis, or interruption of steps in spinosyn biosynthesis. The modified spinosyns may be a new insect control agent or serve as substrates for further chemical modification and the creation of new semi-synthetic spinosyns. The genes are also useful to isolate similar sequences from *S. spinosa* or other species by hybridization.

Sequence 2152 AA;

Query Match 31.5%; Score 2273; DB 20; Length 2152;
Best Local Similarity 32.7%; Pred. No. 5,3e-172;
Matches 600; Conservative 199; Mismatches 542; Indels 496; Gaps 32;

QY 11 EDPPIVAGASCRIPGVIDLSGFMTLLSGSRPTGVRPAER-WDAAMFPDPPAPGTP 69
DB 36 DDPPIVAMSCRYPQGVSSPEDLWQAAAGVADISEVPEDRGMDLDGVFPSPDPSGTSY 95
QY 70 VTRASFLSDVACFDASFGISPREALRMDPAHRLILEVCWEALENAAIPASLVGTETGV 129
DB 96 ACAGGFLQGVSEFDGFGISPREALMDPQGRLLLEVAMEVFERAGLEGQSTRSRGV 155
QY 130 FIGIPSEYEALPATASAEIDAHGIGTMPGVAGRI SYALGRCPCVADVTAYSSSL 189
DB 156 FVGITGODVYASWL--RTPEPAVAGHVLTCGAAAVISGRVAVSFGEGPAVTVDTACSSSL 213
QY 190 VAVHLACSLSGECSTLALAGVSLMSPSTLVMLSKTRALARDRCRFASEADPFRG 249
DB 214 VALHLAAGALRAGECDLALAGVTVWSTPKVLEFSRQRLAPDRCKSFAGADGTWVG 273
QY 250 EGCAVAVVLRKLSGABADGRI LAIVIRGSAINHAGSSGLTVNGSSOEIVLKRALADAGC 309
DB 274 EGAGILLERLSDARNGHEVLAIVRGSAVNDGASNGLITAPNGSSQORVITQALASGL 333
QY 310 AASVGVYEAHGTGTTGDPIEIQALNVAVYGLGRDVPATPLLIGVKTMLGHEPVASGITG 369
DB 334 SVSDVDAVEAHGTGRLDPIEQALITVYGRDRDPGRPLMLGVKSNIGHTQAAGVAG 393
QY 370 LKVLVLSIQHGOIPAHILAQALNPRIISMGDLRTYTRARTPPMDTTRRRACVSGFMSG 429
DB 394 VIKWMARHGOPLPTLHVSESPSPVDWSAGTVQTLTENTPMPRSGRVARVGVSSFGISG 453
QY 430 TNAHVLEAPATCTPPAP-----ERPAILVLSARTASALDQAARLRLDHLTYSQ 483
DB 454 TNAHVILTPQCPVPSQASGPGSGVVDVVPVPMWMSGKTPRLSLQAALATMLTYLDERDV 513
QY 484 CLGDVAFSLATTRSAHEHRLAVATSRGLRALDAAAGQTSPGA VNSIADSSRGLAF 543
DB 514 SSLDVGYSIALTRSLDERAAVVLGSDRETLLCGVVALSAGHSAAGSLVTVGAG-GRIGF 572
QY 544 LFTGGAGTTLGWRGLYDVWSAFREAFD-LCYRL---FNQGLDRLLREVMMAEPASVDA 599
DB 573 VFSGGGGLWGLKGRGLYRAFPVFAAFADEACALDAHIGQETG--VREV---SGSDAQ 626

QY 600 LLDOTAFOPALFTFEVALAALMRSGVEPELVAGHSIGELVAAACVAGFSLDPAVLVA 659
DB 627 LLDRTLMQSGSLFALQVGLLTLDSWGRPEPVILGSHSGELIAAFAAGVSLSGAARLVA 686
QY 660 ARGTLMOALPAGGMVSTIEAEADVAAVAPHAASVSTIAANADQVITAGAGPVHIA 719
DB 687 GRALMOALPFGGGLANPAGBELLWLSLADQGRVGLAANAAGSVVLSGDRVLDLDA 746
QY 720 AAMARGARTALVHSAFHSPMLAPMLAEGRVAESVYSRPSIVLVSNSLGA-CTDE 778
DB 747 GRLOGGIRKMWLVASHAFHSYRMDPMLAEFAELARTVDYRCCEVPITVITGLDDAGR 806
QY 779 VSSPGVYWRHAREVVRPADGVKALAAAGCTFVEVGPSTLLGLVPACMPDA---RPAL 834
DB 807 MSGPDVYWRQVREPVRFADVQALVEHDVATVVEIGFGALSLALIQECVAASDHAGRISA 866
QY 835 LASSRAGDEPATVLEALGGLMANVGLVSNAGLPP-SGGRVPLPTPMQERV----- 888
DB 867 VPAMRRNDEAQKWTALAHVHVGAVDMSFPAGTGAKQIELPTVAFORQRYLVPSD 926
QY 889 ----- 888
QY 927 SGDVTCAGLAGAEHPLLAGVVPVAGGDEVLTGRISVTRTHPWLAEHRYLGEVITAGTALL 986
DB 889 -----IDTRADDAAR 898
QY 987 EIALHAGERLGCERVEBELTLEAPLVLPERRGAIQVQLVAGAPENSGRRPMALYSRPEGAAE 1046
QY 899 GD-----RRAPAGH----- 908
DB 1047 HDWTRHATGRILAPRGENAAGLDADWPAPGALPVLDDEFYRDIAELGLEYGPIFOGLKAAW 1106
QY 909 -----DEV-----BEG----- 915
DB 1107 RQGBEVYAEALPGBTEBSGFEVHPALLDALAHATAVDMDDARLPQWEGVSLAKAAPA 1166
QY 916 -----AVRGDRRSA-----RLDHP- 930
DB 1167 LRVAVPAGDADKSLVCDGTGRFVIVSVDRVLVLSAARTRGARQAHQARLYLSMPTV 1226
QY 931 -----PREGREKV 940
DB 1227 QLPISAPPSCVLLGTSEVSADIGVYDRLSLTAALDAGAPPGVITAPTPGGGRADV 1286
QY 941 E----- 941
DB 1287 RETRHALDLVQWLSQORLINESRLVLTQCAVAVEGPEVTDLAQAALMGLSLRSTQTEH 1346
QY 942 -----AACD-----RPF 949
DB 1347 PDRFVLVDVPPADLLPALPGVLAAGEPOLARRGGAHAPRLAAGSDVDLPVFDGTGWR 1406
QY 950 LEIDEPGLDLVLVITERRAP---GLGEVEIAVDAAGLSFNDVOLAGVPPDLLPGKPN 1006
DB 1407 LEATRPGLDLAL-VDEPTATLGDGEVYIARRAAGVNERDMLIALGMVPG----- 1458
QY 1007 PELLGEGCAGRIYAVEGVNGLVVGQVYIALSAGAFATHTTSAALVLPPOALSIAEA 1066
DB 1459 -VASLGSBGACVVVETGEGVGLAPGDRVMKMTPKAGPLAVADHRVTRIPAGMSFARA 1517
QY 1067 AAMPVAYLTAMVALDRIARLQGERVLIHAATGVGVLAAVQAQVGAENATAGTEKR 1126
DB 1518 ASVPIVELTAYVALVDLAGLRGSSLLVHSAAGCVMAAIDLAHLGEEVATASBQWQ 1577
QY 1127 AYLESGLRVYSDRSRPFVADVAMTGGBGVDVYVLSLSELIDKSFNLRSHGRFVEL 1186
DB 1578 AV--ELSRHILASRTCDPEQOFLGATGCGRVDDVNLNSLAGEPADASLRMLPRGGRFLEL 1635
QY 1187 GKRCVADNOLG-LRPFILNLSFSLVDRGMLEPRPARVALLEBELGLTAAAGFTPPPI 1245
DB 1636 KXTIVRDPVEYADAHAPGVSTOAFDTE-----AGPRIGEMLEHVLBEGRYLEPLPV 1689
QY 1246 ATLPIARVADAFRSMAQAHGLKLVLTLDGPEVOIRIPTHAGAP-----STGDRDLDR 1300

Db 1690 TANDVQAPEALRHLSQAHVGVLTMP-----PMDAAGTVLVGGRG----- 1734
Qy 1301 LASAPAPAPAAALAEFLRTQVSQVLTPEIKYGAE--ALFTRLGMD-SLMAVELRNRIA 1357
Db 1735 -ALGAEVAHVLIERGVNMLVLSRRGPAASGAELVAQLTAVGAESVLOACVADR-ET 1792
Qy 1358 SLTKLSTSTFLSTSPNIALLAQNLDAALATLSLEEV 1394
Db 1793 LAKVLASIPDEHPLTAHVAAVLDGVSSESLTEVL 1829
RESULT 11
AAB70966
ID AAB70966 standard; Protein; 2152 AA.
AC AAB70966;
DT 28-AUG-2001 (first entry)
DE S. spinosa protein fragment encoded by ORF19, SEQ ID 44.
XX Forosamine; trimethylthamose; polyketide synthase; biosynthesis;
XX spinosyn; polyketide aglycone; transgenic plant; insect resistance;
XX macrolide; insecticidal; polyketide synthase.
OS Saccharopolyspora spinosa.
PN DE19957268-A1.
XX 08-MAR-2001.
PD 29-NOV-1999; 99DE-1057268.
PR 27-AUG-1999; 99DE-1040596.
XX (FARB) BAYER AG.
PI Eberz G, Moehle V, Froede R, Velten R, Salas JA;
XX WPI; 2001-267102/28.
DR N-PSDB; AAF88336.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
recombinant production of insecticidal spinosyns and their derivatives
XX
XX
XX Claim 57; Page 200-209; 354pp; German.
XX
XX This invention describes a novel method nucleic acid (I) and its encoded
XX polypeptide (II) containing at least one region that encodes an enzymatic
XX activity involved in biosynthesis of spinosyns. (I) are used (i) to
XX identify, inactive or modulate genes involved in the biosynthesis of
XX (II); (ii) to generate a library of polyketide synthases; (iii) for
XX adding forosamine or trimethylthamose to a spinosyn or polyketide
XX aglycone; and (iv) for recombinant production of the corresponding
XX enzymes, which are used for production of (II), their precursors or
XX derivatives, including production of transgenic plants that express (II)
XX and thus have increased resistance to insects. (I) are also useful as
XX markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
XX macrolides with insecticidal, but not antibacterial, activity, and can
XX also be used to raise specific antibodies, useful for identifying
XX expression clones in a gene bank. Cells transformed with (I) may produce
XX (II) at significantly increased levels or produce new derivatives of
XX (II). This sequence represents an S. spinosa polyketide synthase.
SQ Sequence 2152 AA:
Query Match 31.5%; Score 2273; DB 22; Length 2152;
Best Local Similarity 32.7%; Freq. No. 5.3e-172;
Matches .600; Conservative 199; Mismatches 542; Indels 496; Gaps 32;
Qy 11 EDP1AIVGASCRPLPGVIVLDSGFMTLLSGSRDVTGVAPAR-WDAAMAFDPDDAGKTP 69

Db 36 DDP1AIVMSCRVPEQVSSPEDIMQLAGVDALISEVGDGMDDGVFVDSRPRGTSY 95
Qy 70 VTRASFISDVACFDASFFGISPREALRNDPAHRLILEVCWDALENAIAPALVGTETGV 129
Db 96 ACAGGFLQGSBEFDGFFGISPREALRNDPOORLLLEVAWEFERAGLEORSTRGSRGV 155
Qy 130 FGTGFSREYEAALPQATASAEIDAHGIGTMSVAGISYALRGCVAVDTYSSSL 189
Db 156 FVGTNGQDYASWL--RTPPPAVAGHVLTGGAALVAGVAISFGEGRAVVDTRACSSSL 213
Qy 190 VAVHLACOSLRSGECSTALAGVSLMSLSPSTLVWLSKTRALARDGRCFAFSAEADFGRG 249
Db 214 VALHLAQAALRAGEBDLALAGVYTMSTPKVFLFESSRGLAPDRCKSFPAAGDGTG 273
Qy 250 EGCNAVVIKRLSGAPADGDRILAVIRGSAIINHDAASGLTVPNSSQETVLKRALADGC 309
Db 274 ECAGLLLERLSDARNHSEVLAVVRSANVDGASNGLTAPNGSSQORVITQALASAGL 333
Qy 310 AASSVGYEAHGTGTTGDDPIEQALNAVYGLGRDVATPLIGSVKTMHGHEVYASGITG 369
Db 334 SVSDVAVEAGTGTRLDPIEAQNLATYGRDRDPCGRPLWGSVKSNIHTQALAGVAG 393
Qy 370 ILKVVLSLHGGIIPAHHLAQAALNPRISSGDLRLTVTRARTPMPDMNTPRRAGVSSFGMSG 429
Db 394 VIKMMAMRHQGLPRTLIVESPSPEVDSAGTVQULTENTMPRSGRVRARVGSFGISG 453
Qy 430 TNAHVLEBAPATCTPPAP-----ERPABLVLISARTASALDAQARLDHLETTYSQ 483
Db 454 TNAHVILEQPPGVPSQASPGSGSVVDVPPVPMWMSGTTPALSKQAATMTYTDERDV 513
Qy 484 CUGDVAFLATTRSMEHRLVAATSRGLBAALDAQSGTSPEAVSISADSSGKXAF 543
Db 514 SLDVGYSLALTRSLDRRAVVLGSDRETLGCVALASAGHBSGLVGSVAG--GRIGF 572
Qy 544 LFTGGAGQTLGWGRGLYDWSAFREAFD-LCVRL---FNOELDRPREVMMAEAPASVDA 599
Db 573 VFSGGGGMWGMGRGLYNAFPVFAAFDEACELDAHIGQEIG--VREVV-----SGSDAQ 626
Qy 600 LLDQTAFTQPALFTFEVYALALMSWGVPELVAGHSIGELVACVAGVSLDAVFLVA 659
Db 627 LLDRTLMQSGGLFALQVGLLKLDSWGVPSVVLGHSYGEVLAAPAAVGSISGARLVA 686
Qy 660 ARGRLMOALPAGGAVNSIEAPEADVAALVAAPHAASVSIATAVNAPOQVVIAGAGQPVHIA 719
Db 687 GARLMQALPSCGMALVAPAGEELMSLADQDRVGTAAVNAAGSVVLSGDRVLDLDA 746
Qy 720 AAMARGARITKALHYSHAHSPLMAPMLDARGVASVSYRRPSIVLVSNLSGKA-CTDE 778
Db 747 GRLDGGIRSRWLRSLRSHAFHGYRMDPMLAEFAELARTVDYRCEVPYISTLTGDLDDAGR 806
Qy 779 VSSPGIWRBHAREVVRFDGVKALHAAGAGTFVEYGPSTLTGLVPAQMPDA---RAL 834
Db 807 MSGPYYWRQVREPRFPADGVQALVEHDVATVVEGPGALASALIQEVAASDHAGRLISA 866
Qy 835 LASSRAGDEPARTVLEALGLMAVGLVSMWGLFP--SGGRBRYPLTPYMOBERVY----- 888
Db 867 VPAMRRNDDEAKVNTALAHVHRCGAVDMKSFAGTAKOIEPTVAFOKRIVLVPSD 926
Qy 889 ----- 888
Db 927 SGDVTGAGLAGAHEHLLGAVVPVAGGDEVLLTGRISVTRHPMLAHRVLGEYIVAGTALL 986
Qy 889 ----- 888
Db 987 EIALHAGERLGCERVEELTEAPVLPERGAIQVQLRVGABENSGRRPMALYSREGAAE 1046
Qy 899 GD-----RRAPGAGH----- 908
Db 1047 HDWTHAATGRALPGGEAAGADLADMPAGALPVDLDEFYRLAELGLEYGPIFGSLKAAW 1106
Qy 909 ---DEV-----EEGG----- 915

Dh	1107	RQGEVYAEALPGTEDESGFVHRLDLDAALHATVRNMDDAARLPFQMEGYSLHAKARA	1166
Qy	916	-----AVRGDRSA-----RIDHP-----	930
Dh	1167	LRVEVPPAGDDAKSLVCDGTGTPVIVSDBRLVLSAARRTGARROAHQARLYRLSWPTV	1226
Qy	931	-----PPSGREKV-----	940
Dh	1227	QLPFSACPSCVLLGTSEVSADIQVYDRLSLTALDAGAEPPGVVIAPTPPGGGRADV	1286
Qy	941	B-----	941
Dh	1287	RETRRHLDLVQGMHSDQLNESRLLYTQCAVAVEPEPVTDLAQALMLGLSTQTEH	1346
Qy	942	-----RPF-----RPF-----	949
Dh	1347	PDRFVLVDPEPAQLPALPGVLCGSEFQALRGCAHAPRLAGISDVLVPDGTGWR	1406
Qy	950	LEIBEPGVLDHLVLRVTERAR--GLGEVEIANDAGLSTNDVOLALGMVPDDLPGKPN	1006
Dh	1407	LEATRRGSLDGLAL-VDEPTATAPLGDEEVIAMRAAGVNFDDLIALGMYPG-----	1458
Qy	1007	PLLIGGSCAGRIYAVGEGVNGVLVVGQVILASGAFATHTTSAALVPRQALSIEA	1066
Dh	1459	-VASIGSGAGAVVETGPGVTGLAPRGDVMGNIIPAEFPLAVADHRMTTRIIPAGKSFARA	1517
Qy	1067	AAMPVAVLTWYALDRIRALOPGERVLIHAATGCVGLAAVOMAQOHGAEVHATAGTEPKR	1126
Dh	1518	ASVPVIFLTAYVALVDLAGLRPGESLIVHSAAGCGVMAIQLARHLGAEVATASBDKMQ	1577
Qy	1127	AYLSLGVRYVSDSRDFVADVRAMTGGEGVDVVLNSLGBELDKSFNLLRSGRFVEL	1186
Dh	1578	AV--ELSEHSLASSSTCDPEOQFLGATGARGVDVVLNSLAEFPDASLRMP.PRGGRFPEL	1635
Qy	1187	GKRCVYADNOG-LRPLRNLSFSLVYDLRGMMLEPRPARVRLLBELTGLIAGVFTPPPI	1245
Dh	1636	GKTDRDPEVEADAHPGVSIOAFDTVE-----AGPQIGEMLEHELVLEFGRVLEPPLV	1689
Qy	1246	ATLEPIARVADAFRSMAQOHLGKLVLTLDGDEVOQIRIPTHAGAGP-----STGDRDLDR	1300
Dh	1690	TAMDVROAPBELRLHLSQARHVGKVLVMP-----PVMNAGVVLVTGTS-----	1734
Qy	1301	LASAPAPAPAAALEAFERTOVSQVLRTEIKVGAE--ALFTRLGMD-SLMAVELNRNIEA	1357
Dh	1735	-ALGAEVAVRHVIEIGEVNVLVIVSGRGAAGAEVLVQLTAYGAEVELQACDVADR-ET	1792
Qy	1358	SLKUKLSTFTSTSPNIIALLQNLIDALATLSLERV	1394
Dh	1793	LAKVIASTPDEHPRLTAAVVAAGVLDVGGSESLYVERL	1829

RESULT 12	
AAE10142	
ID	AAE10142 standard; Protein; 7068 AA.
XX	
XX	AAE10142;
XX	
DT	29-NOV-2001 (first entry)
XX	
DE	Streptomyces noursei nystatin gene, NysI (partial).
XX	
KW	Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW	antifungal; antibiotic; PKS type I.
XX	
OS	Streptomyces noursei.
XX	
FH	Key
FH	Location/Qualifiers
FT	Domain
FT	34..448
FT	/label= KS9 domain
FT	/note="ketosynthase (KS) domain"
FT	572..890
FT	/label= AT9 domain
FT	/note="Acyltransferase (AT) domain"

FT		Dmainn	904..1123 /label= DH9_domain
FT		/note= "Dehydratase (DH) domain"	
FT		Domain	1443..1686 /label= KR9_domain
FT		/note= "Ketoreductase (KR) domain"	
FT		Domain	1720..1793 /label= ACP9_domain
FT		/note= "Acyl carrier protein (ACP) domain"	
FT		Domain	1813../2236 /label= KS10_domain
FT		/note= "Ketosynthase (KS) domain"	
FT		Domain	2346..2664 /label= AT10_domain
FT		/note= "Acyltransferase (AT) domain"	
FT		Domain	2678..2890 /label= DH10_domain
FT		/note= "dehydrataee (DH) domain"	
FT		Domain	2993..3229 /label= KR10_domain
FT		/note= "Ketoreductase (KR) domain"	
FT		Domain	3266..3339 /label= ACP10_domain
FT		/note= "acyl carrier protein (ACP) domain"	
FT		Domain	3358..3780 /label= KS11_domain
FT		/note= "ketosynthase (KS) domain"	
FT		Domain	3898..4217 /label= AT11_domain
FT		/note= "acetyltransferrase (AT) domain"	
FT		Domain	4231..4432 /label= DH11_domain
FT		/note= "Dehydratase (DH) domain"	
FT		Domain	4523..4770 /label= KR11_domain
FT		/note= "ketoreductase (KR) domain"	
FT		Domain	4806..4879 /label= ACP11_domain
FT		/note= "Acyl carrier protein (ACP) domain"	
FT		Domain	4801..5335 /label= KS12_domain
FT		/note= "ketosynthase (KS) domain"	
FT		Domain	5432..5754 /label= AT12_domain
FT		/note= "Acylitransferase (AT) domain"	
FT		Domain	5768..5977 /label= DH12_domain
FT		/note= "Dehydrtarase (DH) domain"	
FT		Domain	6068..6315 /label= KR12_domain
FT		/note= "Ketoreductase (KR) domain"	
FT		Domain	6348..6421 /label= ACP12_domain
FT		/note= "Acyl carrier protein (ACP) domain"	
FT		Domain	6454..6873 /label= KS13_domain
FT		/note= "ketosynthase (KS) domain"	
PX		MQ200159126-A2.	
PX		16- AUG-2001.	
PX		08-FEB-2001; 2001MO-GB05059.	
PX		08-FEB-2000; 2000GB-0002840.	
PR		10-APR-2000; 2000GB-0008786.	
PR		14-APR-2000; 2000GB-0009387.	
PA	(UNO-) UNIV NORGE TEKNIKSK NATURVITENSKAPELIGE.		
PA	(SWTF.) SINTEF STIFTELSEN IND TEK FORSK.		
PA	(ALPH-) ALFAARMA AS.		
XX	(SINV-) SINEVENT AS.		
PA	(DZIE/) DZIEGLEWSKA H.		

XX Streptomyces noursei nystatin gene, NysI (complete).
DE Polypeptide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KM antifungal; antibiotic; PKS type I.
XX Streptomyces noursei.
OS
FH Key
FH Location/Qualifiers
FT 41..464
FT /label= KS15 domain
FT /note= "Ketosynthase (KS) domain"
FT 578..889
FT /label= AT15 domain
FT /note= "Acyltransferase (AT) domain"
FT 903..1102
FT /label= DH15 domain
FT /note= "Dehydratase (DH) domain"
FT 1446..1731
FT /label= ER15 domain
FT /note= "Enoylreductase (ER) domain"
FT 1740..1988
FT /label= KR15 domain
FT /note= "Ketoreductase (KR) domain"
FT 2023..2096
FT /label= ACP15 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 2117..2538
FT /label= KS16 domain
FT /note= "Ketosynthase (KS) domain"
FT 2635..2953
FT /label= AT16 domain
FT /note= "Acyltransferase (AT) domain"
FT 2967..3167
FT /label= DH16 domain
FT /note= "Dehydratase (DH) domain"
FT 3257..3500
FT /label= KR16 domain
FT /note= "Ketoreductase (KR) domain"
FT 3539..3612
FT /label= ACP16 domain
FT /note= "Acyl carrier protein (ACP) domain"
FT 3634..4057
FT /label= KS17 domain
FT /note= "Ketosynthase (KS) domain"
FT 4153..4472
FT /label= AT17 domain
FT /note= "Acyltransferase (AT) domain"
FT 4486..4725
FT /label= DH17 domain
FT /note= "Dehydratase (DH) domain"
FT 4997..5245
FT /label= KR17 domain
FT /note= "Ketoreductase (KR) domain"
FT 5277..5350
FT /label= ACP17 domain
FT /note= "Acyl carrier protein (ACP) domain"
XX
PN WO200159126-A2.
XX
PD 16-AUG-2001.
XX
PF 08-FEB-2001; 2001WO-GB00509.
XX
PR 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA (SNTF-) SINTEF STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEWSKA H.

PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVRIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valia S, Ellingsen TE, Sletta H, Gulliksen O;
XX WPI: 2001-557614/62.
DR N-PSDB; AAD17186.
XX
PS Claim 15; Page 255-260; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polypeptide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a PKS type I encoding Streptomyces noursei
CC nystatin gene, NysI (complete).
CC
XX
SQ Sequence 9477 AA;
Query Match 31.4%; Score 2260.5; DB 22; Length 9477;
Best Local Similarity 38.0%; Pred. No. 5.5e-170; Indels 287; Gaps 39;
Matches 614; Conservative 170; Mismatches 544;
QY 10 AEDPIAVGASCLPGGVIDLSGFMTLESGSRDVGKPAER-WDAAMFDPDPAKGT 68
DB 4899 ADPIAVGMSCHYPGGVGASBDMRLALDEVDALISGFADRGMDAGLVDPDRGRT 4958
QY 69 PVTBASFLSVACPDASFPGISPREALMPARLLIEVCEALENNAIPSLVETG 128
DB 4959 YSVQGFRLDVAFFDPFGFISPREALSMPOORLLIETMEAFEHAGIDPVGGRGRTG 5018
QY 129 VFIGIGPSEYEALPQATASABIDAHGGLTWPSVGARISYVALGAPCAVADTA5SS 186
DB 5019 TFWGASTQDYASGVPSSEGS---EGHMITGLTSSVLSGRVSYLFGEFGPAVTLDTA5SS 5075
QY 189 LVAVHLACSLRSGECSTALAGVSLMTSPSTLVMSKTRALARDGRCKAPSAEDGFCR 248
DB 5076 LVAMHLACSLRNGESSLALAGVSLMTSPSTLVMSKTRALARDGRCKAPSAEDGFCR 5135
QY 249 GEGCAVYVTKRLSGARADGRITLAVIRGSAINHDA5SGITVPSGSGEIVLKRALDAG 308
DB 5136 AEGVGLVLERLSDPARANGHQLAVIRGSAVNDGASNGLTAPNGPQORVIRQALANSA 5195
QY 309 CAASVGVVEAHGNGTTLGDPPIEIOALNAVYGGRDVATPLLGSVKTNGHPYASGIT 368
DB 5196 VAPGDIVLEGHGTGLAGPIEIOALNAVYGGRDVATPLLGSVKTNGHPYASGIT 5255
QY 369 GLIKVYLSLQHQIIPALHQAOLNPRI5WGDRLTVTRARTPMDWNTPRRAGV5SFGMS 428
DB 5256 SVIKLVRALQEGVVP5SLHIDRSTHVD5SGAIGLTERTPMPETPRRAA5SFGIS 5315
QY 429 GTNAHVLEBPAPATCTPPAPERPAILV--LSARTASALDAQAARLDHLETP5QCL 485
DB 5316 GTNVHTLIEQAPADEAPTPA--DEPRDGLVPLL5GRGEALRAQAALFLFVERPRPAIL 5374
QY 486 GDVAFSLATRSAMEHRLA5AAT5REGLRAALAAAGCQ5SPGAVR5IA5SSRGKLAFLF 545
DB 5375 TDLAHSLAT5RAL5ERAA5YIAD5RDTLT5GLRAL5SGRDP5GLVQ5TA--GKR5T5FLF 5432
QY 546 TGQGAOTL5GRL5LYDV5SA5FREAF5DL5CVL5FQCELD5PL5REV5MA5PASVD5AAL5DOTA 605
DB 5433 TGQGSQ5RPG5GRL5HRY5PV5FAD5DEV5LARLD5GDP5R5REV5LFA5P5AB5AAL5DR5G 5492
QY 606 FTQPALFT5FYAL5AL5MR5G5VE5PELV5AGH5SIELV5AACV5AGV5L5EDAV5LV5AAG5RLM 665
DB 5493 YQAPALFA5V5VAL5FRILT5SWGL5TPDY5LAGH5SV5ELAA5H5AV5GL5SD5ACT5LV5AAR5RLM 5552

```

OY 666 QALPAGAMVSIIEAPADVAAAIVAPHAASVSIIAVNPADVVITAGAGPYHAIAMAMAR 725
DB 5553 QALPAGAMVVALEAEDEVLPLEGLTDRVSVAANPPSVVAVGEEDVLLADLPLAD 5612
OY 726 GARTYALVSHAFFHSPMLAEPMLAEFGVAESVSYRPSIVLSNLSGKACT-DEVSPPGY 784
DB 5613 GRTYKLVSHAFFHSPMLDMDADPAVARGLTTPPTIPFVSNSGGLATAEQVTRTDY 5672
OY 785 VWRHAREVVRPADGVKALHAAG-AGTFVEVEGPKSTILGLVPACMD-ARPALIASRAGR 842
DB 5673 VVGHVRAVRFADGIDWLATQGVHTFLELGPVGLSAMARESLTDPSTRTALLPLTRGDR 5732
OY 843 DEPARVLEALGLMVAVGSLVWAGLPPS-GGRVRLVPTPYMKRREYVYDTKA----- 893
DB 5733 PEEPALVTAVAAMAHAGARVDMWGYFADHGARRTLPTYAQQRERYWPDPTAATSAPPG 5792
OY 894 -----DDAA----- 897
DB 5793 SALDAEFMAVERDVAAALASLDDATVTAMVPALTRMRKRGQTELDKRWRYTWK 5852
OY 898 -RGDRRAPGA-----GHDEVEGGAVRGDRRSA-----RLDHPPESSG--R 936
DB 5853 PRGATAPALTRGRLVLPVPHDQRODDATAMADVETALGTTVRLTITVTRALA 5912
OY 937 REKVAAGDR-PFRLEIDEPGLDHLVRYTERRAPG-----LGEVEIYVDAAGLSFNDVQ 991
DB 5913 ARITEAADGQGF-----SGVLSLLPLATGAGHPGAPALTTLTITTAVALGDDGIDAP 5966
OY 992 L-----ALG-----WVPDDL-----GKPNPPLLGGECAGRIYA 1021
DB 5967 LWNVTRGAVAVAGRAQVTAPEQAAVWGLGRAVLELPAFGTDLPLTLDGQAAARLRA 6026
OY 1022 VGEVNGLVGQPVITALSAGAFATHTTSAALVLRPQALSAIEAAMPVAVLTAMVALD 1081
DB 6027 VLAATD-----GEDAVALL-----RPSGVFLRLIHAHAPGPTARTAFD 6064
OY 1082 RIARQPERVLIHAATGCVGLAAVQMAOHVGAEN-----ATAGTPKRAVYLSL 1132
DB 6065 PAA-----GTVALITGCTGIGGHVARILRADA-THLLITSRGPAPADALRLLEL 6118
OY 1133 GVR-----YVSDSRS-DRFVADVRAM 1152
DB 6119 GARVTLAACDADADRALAALLAELPDDAPLCVHTTAGVVEHVVDALTPEPFAVLRK 6178
OY 1153 TGGEGVDVILNSLSGELIDKSFNLLRSHGRFVELKRDVCYADNQLGRPFLRN-LSFSLV 1211
DB 6179 TVAAH---HLHELTAADLDAEVLFSSTAGVLGAAGQGVAAANHLDALAEHRRSHGLT 6235
OY 1212 DL-----RGMLEERPARARALLEELGLIAGVFTPPPIATLPIARVADARSMAG 1262
DB 6236 ALSVAMGFWAGSGMAD-----AAELTDRVRGGF--EFLAEP-----AVRALLR 6279
OY 1263 AQHLGKLVLTIGD-----PEVOIRIPTHAGAGPSTGDDLLDR 1300
DB 6280 AIENDDTVALADIMERFORAFVAVRPLPFVADDEPFCRAIPARH-TGAAG---LQQ 6335
OY 1301 LASAAPARAAALEAFRTQVSQV-----RTPEIKVGAELFTRLGMDSLMAVELRNT 1355
DB 6336 LAELPEHRRPAVADLRLRTQVAAVLGHADPRIVE---DDHAFRLDGEFSLITLELRNAL 6391
OY 1356 EASLKLKSTFELSTSPNIALLAQVLDALA-----TALSERVAENLELRGVQ 1404
DB 6392 NNAITGLSIPATLVYDLPTPREWADFLAELLGTLPTDTAATVASTASPKLSHSPF 6446

```

RESULT 14
 AAU10701
 ID AAU10701 standard; Protein; 6095 AA.
 AC AAU10701;
 XX
 DT 12-MAR-2002 (first entry)

```

XX Amino acid sequence encoded by S. cellulosum PKS cosmid pKOS28-26, ORF2.
DE Polypeptide synthase; PKS; catalytic domain; ketosynthase domain;
XX acyl transferase domain; dehydratase domain; ketoreductase domain;
KW acyl carrier protein domain; pKOS28-26.
KM Sorangium cellulosum.
XX
OS
XX Key
FH Location/Qualifiers
FT 34..457
FT /label= Ketosynthase_domain
FT 568..895
FT /label= Acyltransferase_domain
FT 922..1100
FT /label= Dehydratase_domain
FT 1429..1614
FT /label= Ketoreductase_domain
FT 1706..1789
FT /label= Acyl_carrier_protein_domain
FT 1812..2235
FT /label= Ketosynthase_domain
FT 2341..2673
FT /label= Acyltransferase_domain
FT 2978..3162
FT /label= Ketoreductase_domain
FT 3250..3333
FT /label= Acyl_carrier_protein_domain
FT 3356..3781
FT /label= Ketosynthase_domain
FT 3887..4219
FT /label= Acyltransferase_domain
FT 4246..4424
FT /label= Dehydratase_domain
FT 4754..4939
FT /label= Ketoreductase_domain
FT 5031..5114
FT /label= Acyl_carrier_protein_domain
FT 5134..5558
FT /label= Ketosynthase_domain
FT 5664..5991
FT /label= Acyltransferase_domain
FT 6018..6095
FT /label= Dehydratase_domain
XX
XX US6280999-B1.
XX 28-AUG-2001.
XX
XX 31-AUG-1998; 98US-0144085.
XX
XX 22-JAN-1998; 98US-0010809.
XX
XX (KOSA-) KOSAN BIOSCIENCE.
XX
XX Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX WPI; 2001-606536/69.
XX
XX Novel purified, isolated DNA molecule from Sorangium cellulosum having
XX polypeptide open reading frame encoding modules with one or more domains
XX such as ketosynthase, acyl transferase and acyl carrier protein domains
XX
XX Claim 1, Fig 4; 72pp; English.
XX
XX The present invention relates to the isolation of novel Sorangium
XX cellulosum polypeptide synthases (PKS), and the polynucleotide sequences
XX encoding them. The polypeptide synthases include catalytic domains such
XX as ketosynthase domain, acyl transferase domain, dehydratase domain,
XX ketoreductase domain and acyl carrier protein domain. A host cell
XX comprising a PKS ORF (open reading frame) which encodes one or more
XX more PKS domains is useful for producing polypeptide synthases from which

```


AC AAY58577;
 XX 10-APR-2000 (first entry)
 XX Sorangium cellulosum type I polyketide synthase EPOS D.
 DE EPOS D: type I polyketide synthase; epoethione biosynthesis;
 KW polyketide backbone formation; taxol substitute; anticancer.
 XX Sorangium cellulosum.
 OS MO9966028-A2.
 XX 23-DEC-1999.
 PD 16-JUN-1999; 99MO-EP04171.
 PF 18-JUN-1998; 98US-0099504.
 XX 24-SEP-1998; 98US-0101631.
 PR 05-FEB-1999; 98US-0118906.
 XX (NOVS) NOVARTIS AG.
 PA (NOVS) NOVARTIS-BRINDUNGEN VERW GES MBH.
 PI Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
 XX WPI: 2000-097741/08.
 DR N-PSDB: AA255887.
 XX New isolated epoethione synthase genes, used for the recombinant
 PT production of epoethione for use in cancer therapy
 XX Claim 12; Page 139-150; 174pp; English.
 ES This sequence represents a Sorangium cellulosum type I polyketide
 XX synthase, EPOS D, which is one of several epoethione biosynthetic
 CC enzymes encoded by a 68.75 kb contig. Epoethiones A and B are
 CC 16-membered macrocyclic polyketides with an acylcysteine-derived
 CC starter unit; polyketides being synthesised from two-carbon building
 CC blocks, the beta-carbon of which always carries a keto group. Each round
 CC of two-carbon addition is carried out by a complex of enzymes known as
 CC the polyketide synthase in a manner similar to fatty acid biosynthesis.
 CC EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
 CC the thiazole ring formation of epoethiones, and EPOS B, EPOS C, EPOS D
 CC and EPOS E (AAY58575-78) are involved in polyketide backbone
 CC formation. EPO F (AAY58579) is an epoethione macrolactone oxidase, and
 CC the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
 CC involved in transport. Epoethiones mimic the biological activity of
 CC taxol, and may be substituted for taxol in cancer chemotherapeutic
 CC compositions. Epoethiones exhibit a much lower drop in potency against a
 CC multiply drug-resistant cell line compared with taxol, and are
 CC considerably less efficiently exported from such cells by the multidrug
 CC resistance protein (MDR, or P-glycoprotein). Despite the potential of
 CC epoethiones as anticancer agents, they are problematical to produce on a
 CC large scale. Epoethiones are too complex for industrial scale chemical
 CC synthesis, and Sorangium cellulosum is difficult to ferment, producing
 CC poor yields of epoethiones. The nucleic acids of the invention may be
 CC used for the recombinant production of epoethiones in a heterologous host
 CC that is more amenable to fermentation.
 XX
 SQ Sequence 3798 AA;
 Query Match 31.1%; Score 2245.5; DB 21; Length 3798;
 Best Local Similarity 36.2%; Pred. No. 2.1e-169;
 Matches 605; Conservative 191; Mismatches 495; Indels 379; Gaps 43;
 QY 9 AAEADIAIVGASCRPGGVYIDSGFWTLLEGSRDVTGVRPAERWDAAMAFDDPDAPGKT 68
 DB 15159 AADDDIAIVGACRPGGDEGLTYWRHLAGGVNVSTVEPADRMAADWDIDPEVPGRKT 1578
 QY 69 PVTASFLSDVACPDASFFGISPREALRMDPAHRLLEVCWEALNNAAIAPASLVGTETG 128
 DB 1579 YVAKAFPLRDVNSLDAAPFAISPREAMSLDPQORLLLEVSWEAIRAGDDPALRESATG 1638

QY 129 VFIGIPSEYRAALPQATASAEIDAHGGLGTMPSVGAGRIISYALGLRGPCVAVDTPAVSS 188
 DB 1639 VFVEMIGSEHAERQGGDDDDAL-LYGTGMLLSVAAGRISFLGLHGPMTVTDACSS 1697
 QY 189 LVAHLACQSLRSECESTALAGVSLMSPSTLVLSKTRALARDCRKAFAEADGFCR 248
 DB 1698 LVALLHACQSLRSECESTALAGVSLMSPSTLVLSKTRALARDCRKAFAEADGFCR 1757
 QY 249 GEGCAVVVVKRLSGARADGDRILAVTRGSAINHDCASSGLTVPNGSCOEIVUKRALDAG 308
 DB 1758 AEGCAVVVVKRLSGARADGDRILAVTRGSAINHDCASSGLTVPNGSCOEIVUKRALDAG 1817
 QY 309 CAASVGVYEAHAGTGTGDPPIEIOALNAVYGLRDVATPLLIGSKTNLHPEVAGTIT 368
 DB 1818 VAPAEVDFECHGCTGALADPIEVOALGAVYGRGPRAPRPLMLGLVKNLHLEAAAGLA 1877
 QY 369 GLTKVVLSTLOHGOIPAHILHAQALNPRLSWGBDLRLVTTRARTPMDMTPRRAGVSPGMS 428
 DB 1878 GLTKVVLSTLOHGOIPAHILHAQALNPRLSWGBDLRLVTTRARTPMDMTPRRAGVSPGMS 1937
 QY 429 GTNAHVLEEARPAATCTPPAPERPAELIVISARTASALDAQARLRDLHLETFYPSOCLDV 488
 DB 1938 GTNAHVLEEARPAATCTPPAPERPAELIVISARTASALDAQARLRDLHLETFYPSOCLDV 1997
 QY 489 AFSIATTTSAMHRLAVAAATREGIRALDAAAGQTSFGVRSIAD-SSRGKLAFLFTG 547
 DB 1998 AFSIATTTSAMHRLAVAAATREGIRALDAAAGQTSFGVRSIAD-SSRGKLAFLFTG 2057
 QY 548 QGAOTLGKRGGLYDWSAFREAFDLCVRLFNQELDRPLREYVMAEPASVDALDQTAFT 607
 DB 2058 QGOWVWGKRLMAEPIFRALBECORAIDAEAGWSLGLSDEA---ASQGRIDVV 2114
 QY 608 QPALFTFEYALALMRSKGVPELVAGHSISELVAAVGVFSLDEAVFLVAARGELMOA 667
 DB 2115 QPALFTFEYALALMRSKGVPELVAGHSISELVAAVGVFSLDEAVFLVAARGELMOA 2174
 QY 668 LPAAGAMVSIAPAEADVAAVAPRAASVIAVAPADQVYIAGAGOPHAIATAAAMAAAGA 727
 DB 2175 LPAAGAMVSIAPAEADVAAVAPRAASVIAVAPADQVYIAGAGOPHAIATAAAMAAAGA 2234
 QY 728 RTKALHVSASHSPMLAEMLEAFGRVAVESYRRPSIVLVNLSGACTDEVSPPGVWR 787
 DB 2225 FWRQVKVDVAHSQVQVDBELBELIALGAIRPAAAVPMRSTVGVIAGBELASVYAD 2294
 QY 788 HAREVVRPDAVKAHAAAGCTFVEVGKSTLLGLVPCMPDAP--ALLASSRAGDEP 845
 DB 2295 NLRQPVRRPAAQAALLBEGGPALFTIEMSPHILVPLDEIQTAEGCGAAVSLRGGDER 2354
 QY 846 ATVLEALGLMAVGLVSMAGLFPSSGRRVPLPTPMQREERYWIDTKADDAARGRAPG 905
 DB 2355 ATVLEALGLMAVGLVSMAGLFPSSGRRVPLPTPMQREERYWIDTKADDAARGRAPG 2406
 QY 906 AGHDEVEGCAVGGDRRSARLDHP----- 930
 DB 2407 AGHDEVEGCAVGGDRRSARLDHP----- 2457
 QY 931 -----PESGRREKVEAADRPRLEIDFG---VLHVLV-----RTYE----- 967
 DB 2458 GEAVLPSPAAYVEMALAA-----VDLYGTATLVLEQALERALAVSEGRIVQVALS 2510
 QY 968 RRAPLGEVEIAY-DAAGLSFNDVQALAGV-----PDDLPGK 1004
 DB 2511 EBGPERASFOVSREBAGRSM--VRHATGHVCSGSSAVGALKENPMETQRCBSVLSE 2566
 QY 1005 PNPPL-----IG-CECAGRIYAVGCVGLVGGPVIALSAGAFAT 1045
 DB 2569 ALYPLNHALDYGPCFCGVGQVWLGTEGVGRVLPED-----MASSGAVRI 2617
 QY 1046 H-----VTTSAALV-----LPP-----QALSAI----- 1064
 DB 2618 HPALIDACFOVLTALTTPESIEIRRLRLTDLHEPDLPSRASAPVNOAVSDTWLMDAALDGG 2677

QY 1065 --EAAAMPV-----AYLTAWYALDRAR-----LQGERVLIHAATGGVGLA 1104
Db 2678 RROGASVVDVLDVLSFPAKMEWELAQAYIIIGTLRIWNVFCAGERTIDELVRLQIS 2737
QY 1105 AV-----QNAQH-----VGAEVHATAG-----TPEKRAYLESIGVRYVSDSRDRVA 1147
Db 2738 VVYRKVIKRWMEHLVAIGILVGDSEHFVSQPLPEPDLAVALLEAG-----RVFA 2787
QY 1148 DVRA---WT---GGSGVDVVL-----NSLSGELI 1170
Db 2788 DLPVLFKCKEPAGERLADVLGKTLALEILFPGGSFDMARIRYRDSPIARISNGIWRGV 2847
QY 1171 DKSPNLRSHGRF--VELG-----KRDY-----ADNOLGLR 1200
Db 2848 ESAARVVAPEGMFSLIEIGAGTGAATTAAPVLLPDRTEYHFTDVSPLFLARAEGRPDY 2907
QY 1201 PFLRNLSFSLVD-----LRGMLEPARAVRALLEELGLIAGVF-- 1240
Db 2908 PFLK---YGILDVDOEPAGOGYAHQRPDVIIVANVIAHTRDIRATAKRLSLAPGGILV 2964
QY 1241 ---TPPI-----ATLPIARVADAFPSMAQAQHL-----GK 1268
Db 2965 LVBGTGHPIMFDITTTGLIEGWOKYEDDLRIDHPLIPARTWCVDLRRVGFADAVSLPGDS 3024
QY 1269 LVLTLGDPEVQIRIPTHAGAG--PSTGDRDLDRLASAPAPAAA-----A 1312
Db 3025 PAGILGGHVIUSRAAGIAGACDSSGES-----ATESPARAVRQEWADGSADVYHRMA 3078
QY 1313 LEA--FLRTQVSQVLRTPEIVGAEALFTRIGMDSMAVEILNRIEASLK 1360
Db 3079 LERNYFHRPRGQVWVHGRLRTGGGA--FTKALAGDILLFEDTQGVAAEVQ 3127

Search completed: October 2, 2003, 17:33:51
Job time : 133.763 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 37.1179 seconds
(without alignments)
2912.986 Million cell updates/sec

Title: US-10-014-717-8
Perfect score: 2124
Sequence: 1 MTOEANOSETKPAFDKFPKPF.....HPAFNIESLNVILKPSKAG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_RODENT:*
12: SP_VIRUS:*
13: SP_VERTEBRATE:*
14: SP_UNCLASSIFIED:*
15: SP_VIRUS:*
16: SP_BACTERIAP:*
17: SP_ARCHAEP:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2124	100.0	419	2	Q9L8C4
2	2107	99.2	419	2	Q9K1Z4
3	532.5	25.1	395	2	Q8K2M6
4	497	23.4	417	2	Q8R3Z4
5	488	23.0	416	2	Q87605
6	486.5	22.9	415	16	Q92005
7	486	22.9	376	16	Q31785
8	483	22.7	399	2	Q93H43
9	481.5	22.7	418	16	Q9HYR4
10	469	22.1	419	2	Q8KX32
11	459	21.6	416	2	Q8KSY0
12	453.5	21.4	410	2	Q9XSP9
13	451	21.2	415	16	Q8UFY8
14	447.5	21.1	401	2	Q9KHU7
15	446.5	21.0	394	2	Q8KNC8
16	443	20.9	399	2	Q8GGQ1

17	442.5	20.8	434	16	Q9CBE7	Q9CBE7 mycobacteri
18	441	20.8	433	16	Q9F200	Q9F200 streptomyc
19	433.5	20.4	399	2	Q93H81	Q93H81 streptomyc
20	431	20.3	379	2	Q93HFO	Q93HFO streptomyc
21	430	20.2	411	2	Q60005	Q60005 streptomyc
22	430	20.2	417	2	Q59910	Q59910 streptomyc
23	429	20.2	411	16	Q9X8G3	Q9X8G3 streptomyc
24	427.5	20.1	407	16	Q9RJ07	Q9RJ07 streptomyc
25	427.5	20.1	416	2	Q93H80	Q93H80 streptomyc
26	425	20.0	511	2	Q52569	Q52569 amycolatops
27	423	19.9	397	2	Q8KIC8	Q8KIC8 nocardia ae
28	421.5	19.8	444	16	Q91107	Q91107 pseudomonas
29	420	19.8	411	2	Q32460	Q32460 actinomadir
30	417	19.6	407	2	Q59819	Q59819 streptomyc
31	416.5	19.6	398	16	Q8XRX2	Q8XRX2 ralsconia s
32	413	19.4	397	2	Q59523	Q59523 micromonos
33	412	19.4	397	16	Q8PHS9	Q8PHS9 xanthomonas
34	411	19.4	404	2	Q54302	Q54302 streptomyc
35	409.5	19.3	416	16	Q982V9	Q982V9 rhizobium 1
36	408	19.2	404	2	Q93HU0	Q93HU0 streptomyc
37	407.5	19.2	436	2	Q9S4D6	Q9S4D6 streptomyc
38	406	19.1	310	2	Q32927	Q32927 mycobacteri
39	405.5	19.1	418	2	Q8GMF6	Q8GMF6 streptomyc
40	403.5	19.0	420	2	Q92HQ1	Q92HQ1 streptomyc
41	402.5	19.0	406	2	Q87192	Q87192 streptomyc
42	401.5	18.9	397	2	Q8KND6	Q8KND6 micromonos
43	397.5	18.7	416	3	Q96MS9	Q96MS9 trichosporo
44	396.5	18.7	397	16	Q8P6C3	Q8P6C3 xanthomonas
45	393	18.5	399	2	Q93NXL	Q93NXL streptomyc

ALIGNMENTS

RESULT 1	Q9L8C4	PRELIMINARY;	PRT;	419 AA.
AC	Q9L8C4;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Cytochrome P450 167A1.			
GN	CYP167A1 OR EPOF.			
OS	Polyangium cellulosum.			
OC	Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;			
OC	Scorariaceae; Polyangiaceae; Polyangium.			
OX	NCBI_TaxID=56;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SO CE90;			
RX	MEDLINE=20130945; PubMed=10662695;			
RA	Molnar I., Schnupp T., Ono M., Zirkle R.E., Milamow M.,			
RA	Nowak-Thompson B., Engel N., Toupet C., Strittmann A., Cyr D.D.,			
RA	Goriach J., Mayo J.M., Hu A., Goff S., Schmid J., Ligon J.M.;			
RT	"The biosynthetic gene cluster for the microtubule-stabilizing agents			
RT	epoхиlonas A and B from Scoringium cellulosum So ce90."			
RL	Chem. Biol. 7:97-109(2000).			
CC	-I- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.			
DR	EMBL; AF210843; AAF26924.1; -.			
DR	HSSP; P33006; ICPT.			
DR	InterPro; IPR001128; Cytochrome_P450.			
DR	Pfam; PF00067; P450; 1.			
DR	PRINTS; PR00385; P450.			
KW	Heme; Monooxygenase; Oxidoreductase.			
SQ	SEQUENCE 419 AA; 46750 MW; 1355857FB04FB8FE CRC64;			
Query Match	100.0%; Score 2124; DB 2; Length 419;			
Best Local Similarity	100.0%; Pred. No. 8.5e-145;			
Matches	419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 MTOEANOSETKPAFDKFPAPGVADPPPAIERLRKATPIFYWEGSGSWLTRHYDVA 60			
DB	1 MTOEANOSETKPAFDKFPAPGVADPPPAIERLRKATPIFYWEGSGSWLTRHYDVA 60			

```

QY 61 VFDERFAVSREEMESSAESSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
DB 61 VFDERFAVSREEMESSAESSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
QY 121 LTRAEIORTVDQLDARSQGEFDDVRYDAEGI PMRAISALLKVPACDEKFRFGSATA 180
DB 121 LTRAEIORTVDQLDARSQGEFDDVRYDAEGI PMRAISALLKVPACDEKFRFGSATA 180
QY 121 RALGVGLVPOVDEETKTIVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRUS 240
DB 121 RALGVGLVPOVDEETKTIVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRUS 240
QY 181 RALGVGLVPOVDEETKTIVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRUS 240
DB 181 RALGVGLVPOVDEETKTIVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRUS 240
QY 241 TKELVALVGAIIAAGDTTIIYLAFAVNLNLSPEALELVKAPGIMRNALDEVLFNDI 300
DB 241 TKELVALVGAIIAAGDTTIIYLAFAVNLNLSPEALELVKAPGIMRNALDEVLFNDI 300
QY 301 LRIGTVRPARQDLEYCGASIKKGEWFLIPBALRDGVFSRPDVFVDRDTSASLAYGR 360
DB 301 LRIGTVRPARQDLEYCGASIKKGEWFLIPBALRDGVFSRPDVFVDRDTSASLAYGR 360
QY 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPARNIESLNVILKPSKAG 419
DB 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPARNIESLNVILKPSKAG 419

RESULT 2
ID 09K124 PRELIMINARY; PRT; 419 AA.
AC 09K124;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
DE EPOK.
GN EPOK.
OS Polyangium cellulolum.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Soranginae; Polyangiaceae; Polyangium.
OX NCBI_TaxID=56;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RX MEDLINE=20393058; PubMed=10831849;
RA Julien B., Shah S., Ziermann R., Goldman R., Katz L., Khosla C.;
RT "Isolation and characterization of the epoH1 gene biosynthetic gene
cluster from Sorangium cellulolum."
RL Gene 249:153-160 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SMP44;
RX MEDLINE=20115953; PubMed=10649995;
RA Tang L., Shah S., Chung L., Carney J., Katz L., Khosla C., Julien B.;
RT "Cloning and heterologous expression of the epoH1 gene cluster."
RL Science 287:640-642 (2000).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF17189; AAF62866.1; -.
DR HSSP: P33066; 1CPT.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450.
DR PRINTS: PR00385; P450.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 419 AA; 46750 MW; FCCEDAI93563688F CRC64;

Query Match 99.2%; Score 2107; DB 2; Length 419;
Best Local Similarity 99.3%; Pred. No. 1,4e-143;
Matches 416; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

DB 61 VFDERFAVSREEMESSAESSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
QY 121 LTRAEIORTVDQLDARSQGEFDDVRYDAEGI PMRAISALLKVPACDEKFRFGSATA 180
DB 121 LTRAEIORTVDQLDARSQGEFDDVRYDAEGI PMRAISALLKVPACDEKFRFGSATA 180
QY 181 RALGVGLVPOVDEETKTIVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRUS 240
DB 181 RALGVGLVPOVDEETKTIVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRUS 240
QY 241 TKELVALVGAIIAAGDTTIIYLAFAVNLNLSPEALELVKAPGIMRNALDEVLFNDI 300
DB 241 TKELVALVGAIIAAGDTTIIYLAFAVNLNLSPEALELVKAPGIMRNALDEVLFNDI 300
QY 301 LRIGTVRPARQDLEYCGASIKKGEWFLIPBALRDGVFSRPDVFVDRDTSASLAYGR 360
DB 301 LRIGTVRPARQDLEYCGASIKKGEWFLIPBALRDGVFSRPDVFVDRDTSASLAYGR 360
QY 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPARNIESLNVILKPSKAG 419
DB 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVFGYHAPARNIESLNVILKPSKAG 419

RESULT 3
ID 08K2M6 PRELIMINARY; PRT; 395 AA.
AC 08K2M6;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
DT 01-MAR-2002 (TEMBLrel. 23, Last annotation update)
DE Cytochrome P450 enzyme.
GN BLOI.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OK2;
RA Sasaki M., Koriyagi T., Kuruu Y.;
RT "Genetic analysis of biotin operon in Bacillus subtilis natto OK2."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AB088066; BAC03244.1; -.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 395 AA; 44869 MW; 6D4B5C445107D225 CRC64;

Query Match 25.1%; Score 532.5; DB 2; Length 395;
Best Local Similarity 32.1%; Pred. No. 3.3e-30;
Matches 130; Conservative 96; Mismatches 138; Indels 41; Gaps 11;

```


Qy 356 LAYGRGPVCPGVALAEIAVGTIFRRFPKLTETP---VFQHPAFRNIESLNV 411
 Db 345 LARGHGHFCIGAPLARLEARIIVRALRECPDLADVSGELVWYENPMIRGLKALPI 403

RESULT 6

092005 PRELIMINARY; PRT; 415 AA.
 AC 092005;
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Putative cytochrome P450-like monooxygenase protein
 DE (EC 1.14.14.1).
 GN R01252 OR SMC01812.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxId=382;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Bailly-Hubier F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boissard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Maury D.,
 RA Pohl T., Portetle D., Puhler A., Purnelle B., Ransperger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591786; CAC45831.1; -;
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Complete proteome.
 SQ Oxidoreductase; 415 AA; 46750 MW; 8696BA61B71B4E7 CRC64;

Query Match 22.9%; Score 486.5; DB 16; Length 415;
 Best Local Similarity 31.6%; Pred. NO. 7.2e-27;
 Matches 127; Conservative 77; Mismatches 179; Indels 17; Gaps 10;
 Qy 22 PGYAEDEPPAIEERLEATPIFYWDEGRSWLTRYHDVSAVFRDERFA-----VSREEMES 76
 Db 23 PRFRNPDLPAVYALAHQCPAFWBEPOQWFAEGVNSLDRRFRGRLVHATREELG 82
 Qy 77 SAEYSSAIPELSDMKKYGFLGPPEDHARVKLVNPSFTSRADLDLRAEIQRTVDLLDA 136
 Db 83 MPEPKPHLKDFDLAEHSLLEBPATRTLRITLVNRAVFSROIELEPEIEALSHAVIDG 142
 Qy 137 RSQGEEDVYRDYAEIGPMAISALILKVPACDEKFRFQSAVARALVGLVPOVDETK 196
 Db 143 FEKDETELEKTYAETIPVTITARMGIPVEAARLIDMSHRMVKVFN--DSLETFD 200
 Qy 197 TLVAVTEGIALHDVDERRRNPLENDVLTMLQAEADSRSLSTKEVALVAGAIAGT 256
 Db 201 ANNAS-AEFDVYKGIARKTNPAD-DLTHMITSKOGBRISDDELISTVLLNAGH 258
 Qy 257 DTTIYLAFAVNLNLSF-EALELVKAEPLMRNALDEVLPDNILIRIGTVRARQDLEY 315
 Db 259 EATVHQIGNAVRTILSGSLPAELFSDKATER-TVEECRFAAPLHI-FQRYALMDIEL 316
 Qy 316 -CGASTIKGMMVLELPSALRDGTVERPVPVDRPRTGSLAVGRGHPVCSVLARLE 374
 Db 317 ENGIALKKGKIGIMGAANVDPKRFSSPTFRPDNRGANVSGAGLHFCIGAPLARLE 376
 Qy 375 AEIAYGTIFRRFPKLTETP---VFQHPAFRNIESLNV 412
 Db 377 LQISLPILFRRLPGMRKNEPVPYDAFH--FHGIERLDIV 414

RESULT 7

031785

ID 031785 PRELIMINARY; PRT; 376 AA.
 AC 031785;
 DT 01-JAN-1998 (TEMBLrel. 05, Created)
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Putative cytochrome P450 107K1.
 DE
 GN PKS OR CYP107K1.
 OS Bacillus firmicutes.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxId=1423;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertolo M.G., Bessières P., Bolotin A., Borcherdt S.,
 RA Bortiss R., Boursier L., Brans A., Brun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conterton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Eutian K.D., Errington J., Fabret C., Fertari E., Foulger D.,
 RA Fritsch C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Chim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Harech U., Harwood C.R., Henauc A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
 RA Pirescan E., Puig P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
 RA Sekiguchi U., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vasseroiti A.,
 RA Viari A., Wambut R., Wedler H., Welzenegger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoko K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein B., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; Z99113; CAB13607.1; -;
 DR HSSP; Q00441; 10XA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Heme; Hypothetical protein; Monooxygenase; Oxidoreductase;
 KW Complete proteome.
 SQ SEQUENCE 376 AA; 43418 MW; C3899600FFD87ECD CRC64;

Query Match 22.9%; Score 486; DB 16; Length 376;
 Best Local Similarity 31.6%; Pred. NO. 6.8e-27;
 Matches 120; Conservative 76; Mismatches 156; Indels 26; Gaps 7;
 Qy 17 EKPAPGVAEDPPAIEERLEATPIFYWDEGR-----SWLTRYHDVSAVFRDERFA-- 68
 Db 8 FHPHGKFPNNPSVLRFRREPEIRHFLKFGATYPAWLITRYDDCAAFKDNRTID 67
 Qy 69 ---VSREEMESAEYSSAIPELSDMKKYGFLGPPEDHARVKLVNPSFTSRADLDLRAE 125
 Db 68 VKNVNMQEOIKMLNVSEIDIDFVSD---HMLAKOTPDHRLKSLVHQATPTTLENLRS 123

```

QY 126 IRTVTDQLDARSQGEEDVVDVAEGIMRAISALLKVPACDEKFRFGSATAAL-- 183
DB 124 IEQIQLDLEMEKENKADIMKSPASPLPIYISELMJPKEDRSQFOIWTNAMDTSBG 183
QY 164 GVGGLVPQVDEEKTIVASTBEGIALHVDLDERRRNPLENDVLTMLQAEAGSRISTGE 243
DB 164 NRELNVQALREFKDYIAK-----LTHD---RRKP-KDOLI SKLVHAENGSKLSEKE 232
QY 244 LVALGATIAAGTDTTITLYIAFAVNLILSRPALIELVKAEPGLMRNALDEVLPFNILRI 303
DB 223 LYSMLFLVLVALETTVNLGSGTLLALQHKKECECTKQPEMIATVAVELLRTSPVVM 292
QY 304 GTVPARQDLEYCGASIKKGEVFLIIPALRDGVTFSPDVPVARDTGSALVGRGPH 363
DB 223 MANRMAIEDFTYKGSIKRQDMIFIGISANDPNEFENPEITINRSPNRIISFGFIH 352
QY 364 VCPGVSLARLEAIVGTIF 383
DB 353 FCLGAPLARLEBGIHAFKAAF 372

```

RESULT 8

```

QY 093HA3 PRELIMINARY; PRT; 399 AA.
AC 093HA3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Cytochrome P450.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=33903;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: Deducing the ability of producing secondary
RT metabolites."
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AB070947; BAB69286.1; -.
DR HSSP; P23395; 1JFB.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 399 AA; 44538 MW; CBOD90378008F99D CRC64;

```

```

Query Match 22.7%; Score 483; DB 2; Length 399;
Best Local Similarity 32.2%; Pred. No. 1.2e-26;
Matches 126; Conservative 66; Mismatches 163; Indels 36; Gaps 7;

```

```

QY 24 YAEDEPPIERLREATPIFYWDEGR---SWVLRHYDVSAVRDRFAVSRE---EMES 76
DB 3 FGADVPPEYAMLRABEPRQVLEGKGLGLVTRKIEDVRKLSIDPRMSADPNNAIDVQOE 62
QY 77 SAEYSSAIPELSDMKKYG--LFGLPEDHARVRLKVNPSFTSRAIDLRAEIQRTVDQL 134
DB 63 AKGR-----PLEDRGLGTHLLTDAPEHTRRLVSTAFTRARVGLRAQVQHTDGL 118
QY 135 DARSQGEEDVVDVAEGIMRAISALLKVPACDEKFRFGSATAALVGLVQVDE 194
DB 119 DTIVRGQAEELGDFAPLAITVICELVGPKADDPVFQWTKDFFR-----W 166
QY 195 TKTLVASVTEG-----LALHADVDERRRNPLENDVLTMLQAEAGSRISTKE 244
DB 167 TINTSDAQDRGQARVGLRDLLELTRVLVDRKRQPADG-LVDALLIARDDDRINAEVL 225
QY 245 VALVGAIIAAGTDTTITLYIAFAVNLILSRPALIELVKAEPGLMRNALDEVLPFNILRIG 304

```

```

DB 226 LSMMSILLVGFEFTVNLIGNGTLALLRHPPQALLRRRPELVDSALEMRLDYDSFETA 285
QY 305 TVRFARQDLEYCGASIKKGEVFLIIPALRDGVTFSPDVPVARDTGSALVGRGPH 364
DB 286 TWRFPLEPIEVAAGRIEKGHPVLJLSANSDGAKFAFPDPDFVTRADPAHVAFGAHF 345
QY 365 CPGVSLARLEAIVGTIFRRFPPEKLETP 395
DB 346 CLGAPLARLEBGIHAFHGLRLRLPGIALSVPP 376

```

RESULT 9

```

QY 09HYR4 PRELIMINARY; PRT; 418 AA.
AC 09HYR4;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Putative cytochrome P450 107S1.
OS CYP107S1 OR PA3331.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN 11
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westhede-Wedman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Iarbig X., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RT Nature 406:959-964(2000).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AE004755; AAG06719.1; -.
DR HSSP; Q00441; 1OXA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
KW Heme; Hypothetical protein; Monooxygenase; Oxidoreductase;
KW Complete proteome.
SQ SEQUENCE 418 AA; 8231E91DB10F351 CRC64;

```

```

Query Match 22.7%; Score 481.5; DB 16; Length 418;
Best Local Similarity 33.5%; Pred. No. 1.7e-26;
Matches 116; Conservative 73; Mismatches 140; Indels 17; Gaps 5;

```

```

QY 50 WVLTRYHDSAVFDERARVREMESSAEYS---SAIPLSDMKKYGLFGLPPEDHAR 105
DB 41 WVVTRYRRARKVRLNHP--GVRRDARQAELVAKRTGSPRAGIGEGLSHMLNLPDPDHR 98
QY 106 VRKLVNPSFTSRAIDLRAEIQRTVDQLDARSQGEEDVVDVAEGIPMAISALLKVP 165
DB 99 LRSLVGRAFTROVERLDPHIERITTEALLDMAGREODLAADPAIPITIAVIFELGIP 158
QY 166 AECEKFRFGSATAARALVGLVQVDEETTLVASTBEGIALHVDLDERRRNPLENDV 225
DB 159 EAEREHAQSNRQAEELLS-----PEAQALDAQVDYRLV---LEAKRRPAD-DV 207
QY 226 LTMLOAEADSRISTKEVALVGAIIAAGTDTTITLYIAFAVNLILSRPALIELVKAEPG 285
DB 208 YSGLVQAADDESGQISEALVMAHLMWSGFEITNMIGNALVTLVNPEDQALLRQPE 267
QY 286 LMRNALDEVLPFNILRIGTRFARQDLEYCGASIKKGEVFLIIPALRDGVTFSPRPV 345
DB 268 LIPNMBELVNHDSPPVARSMTKFTVEDVELDGVTIPAGEYITLVNLTNHNDAERFDPDR 327
QY 346 FVVRDGTGASLAVYGRGPHVCPGVSLARLEAIVGTIFRRFPPEKLT 391

```


RA Mao Y.Q., Varoglu M., Sherman D.H.;
RT "Molecular characterization and analysis of the biosynthetic gene
cluster for the anticancer antibiotic mitomycin C from Streptomyces
lavendulae NRRL 2564.";
RL Chem. Biol. 6:251-263(1999).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL: AF127374; AAD28449.1; -.
DR HSSP: Q5080; 1107.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KM Hypothetical protein; Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 410 AA; 45887 MW; 1B9C78D8AD34E748 CRC64;

Query Match 21.4%; Score 453.5; DB 2; Length 410;
Best Local Similarity 32.5%; Pred. No. 1.7e-24;
Matches 134; Conservative 61; Mismatches 140; Indels 77; Gaps 13;

QY 19 PF----APGYADDPAPAIERLRERATPIFYWDE-----GRSWLTRYHDVS 59
DB 12 PFGAGAPAY---PFHAPDRLE---PDYWEPLRERERPLQRYVTLPGGEAMLATRYQDVR 65
QY 60 AIFRDERPVSREBESSAIESSAIPELSDMKYGLFGIPPE-----DHARVKLVN 111
DB 66 AIFRDERPVSREBESSAIESSAIPELSDMKYGLFGIPPE-----DHARVKLVN 112
QY 112 PSFTSAIDLRRAEIORVTDOLLARSQGEF---DVRDVAEGIPMAAISALLKVPAP 167
DB 113 KYTFPRRVDKMRPLQRTADGLDA---MEWGPRAVDLPSLPRAVSMICELLGVPE 169
QY 168 CDEKRRFSGATARALGVLPQVDEETKLVAIVTEGLALHDVLD---ERRRNPEN 223
DB 170 DKRRFCVSDALLTT-----TAHTPAQVRDVMQWHDYIGLVAQRARRPTA- 216
QY 224 DVLTLMLQADGSRISTELVALVGAITIAAGDTTITVIAPAVNLRLSPALIVKAE 283
DB 217 DLIGSLVTRDEDEKLTGELVLAELIAGETSSASQIPNELYLVFHHPOLLEIRND 276
QY 284 PGLMRALDEVRFNDILIRIGTV---REFARODLEVCASIKKGEVNFLLISALRDGV 339
DB 277 HLLIPDAVELRFR---VPIGTVDGPPRTATEDVEGLVVRAGEVTPSMGANNDPRL 333
QY 340 FSRPVDVFRDYGASLAVGRPHVCPGVSLARLEAIVAGTIFRRFPENKL 391
DB 334 FTDPELDLARPNPRLGFGAGPHGHLGQALRVELQITLTLFRYPRRL 385

RESULT 13

Q8UFY8 PRELIMINARY; PRT; 415 AA.
AC Q8UFY8:
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Cytochrome P450 hydroxylase.
GN ATU1256 OR AGR C.2319.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxId=176299;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okuda Y.K., Zhou Y., Chen L., Wood G.B., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Boyce D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McCelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,

RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science 294:2317-2323 (2001).

RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quirillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Hounell K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Mollam C., Allinger M., Doughty D., Scott C., Lappes C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cleio C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL: AE008087; AAL42265.1; -.
DR EMBL: AE008053; AAK87051.1; -.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam: PF00067; P450; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KM Complete proteome.
SQ SEQUENCE 415 AA; 47458 MW; 920520E0C6C95DDA CRC64;

Query Match 21.2%; Score 451; DB 16; Length 415;
Best Local Similarity 29.1%; Pred. No. 2.6e-24;
Matches 118; Conservative 82; Mismatches 175; Indels 30; Gaps 9;

QY 22 PGYADDPAPAIERLRERATPIFYWDEGRSWLTRYHDVAVFDERP-----AVREKWE 75
DB 24 PAFVNDPNPVYALHAOCPTFYMEORQWFFTCYDHSVSTLLRRFRGQILHVASREE-- 81
QY 76 SAEYSNAIPE-LSMK-----KXGLGFLPEDHARVRKVNPSFTSAIDLRRAEIOR 128
DB 82 -----IGUPELHVHVKFPLDABQSLLEBPETHRLKTLINRAVSHVDMKPELEE 135
QY 129 TVDQLLDARSQGEEDVDVADYAGIEMRAISALLKVPACDEKFRFSGATARALGVGLV 188
DB 136 LANRLIEAFANGFELLSSYADIIPTVMIMAMICIPMEMGQQLKWSHA---YQGMWF 192
QY 189 POVDEETKLVAIVTEGLALHDVLDERRRNPENLVDTMLQADGSRISTKELVALV 248
DB 193 KTPPEDELADLADQAEPSDYYRVAERRAEP-KODLLSHMHTHKQCYLTDDELVSFT 251
QY 249 GATIAAGDTTITVIAPAVNLRLSPALIVKABGLRNALDEVLPNDILIRIGTVRF 308
DB 252 IYLVNAGHATVHQIGNSVRITLESGLDPKTLFHDERTERTVETLRCAPVHI-PQW 310
QY 309 ARODLEYCASIKKGEVNFLLIPSAIRDGTVFSRPDVFVRDYGASLAVGRPHVCPV 368
DB 311 VLEPVEIDGVPFKRGDKVSLIAAANDPAPFSDPLAQPDNNEGANSFPGIHFICGA 370
QY 369 SIARLEAIVAGTIFRRFPENKLKETPVFG--YHPAFNNIESLVN 411
DB 371 PLARLELNLALPLFLFKPLGFKIAEPPVKDYVH--FGLERLDL 413

RESULT 14

Q9KHJ7 PRELIMINARY; PRT; 401 AA.
AC Q9KHJ7:
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Putative cytochrome P450 107R1 (Putative cytochrome P450 monooxygenase
DE Encl).
GN CYP107R1 OR ENCR.
OS Streptomyces maritimus.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=115828;
RN (1)
RP SEQUENCE FROM N.A.

RA Piel J, Hoang K., Moore B.S.;
RA "Natural Metabolic Diversity Encoded by the Enteroicin Biosynthesis
RT Gene Cluster."; *J. Am. Chem. Soc.* 122:5415-5416(2000).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF254925; AAF81737.1; -.
DR HSSP; 000441; 10XA.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; P800385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Heme; Monooxygenase; Oxidoreductase.
SQ SEQUENCE 401 AA; 44184 MW; 73CDA22BD475B2E7 CRC64;

Query Match 21.1%; Score 447.5; DB 2; Length 401;
Best Local Similarity 31.9%; Pred. No. 4.4e-24;
Matches 126; Conservative 70; Pident matches 166; Indels 33; Gaps 12;

QY 9 SETKPAFPKPPAPVADPPPAIRLRLEAPPIF-----YWDGRSWLTTHDOVSAYFR 63
DB 3 THTQQLRPF-PPAPPAELHMEPAFAQLRKEEDISVRLPY--GGEAMLVTRYDITKVLG 59
QY 64 DERFAVSREWESSAESAIPELSDMKYQ-LFGLPEPDHARVAKLVNPSFTSAIDL 122
DB 60 DPRFS-----RAAQHQAPRIQDPDPAEGEGLMSLDPDHTRLKTYAGFTKRVED 113
QY 123 PAEIORIVQDILDA-RSGQEEFDVRYDAEGIPKPAISALLKVPAPCEDEKRRRFGSATAR 181
DB 114 RPATGRIAEILLLEMAEGAPADLVASALPLPVIVICDLLGVDPDDREOLRGWSDAL-- 171
QY 182 ALGVGLVQVDEBTILVASVTEGLATLHDVYDERRRNPLENVDVITWILQA-EADGSRLS 240
DB 172 ---ISTTACTPAESAAAQAMADHPAL--VSQRRQPTD-DLGLALVQWYDREGGLR 224
QY 241 TKEILVALGATIAAGTDTTIVLIAFAVINLILRSPALRLVKAEPGIMENALDEVLRPDI 300
DB 225 DEEVLRLRDLIAHETTASQIANCTYLLQRPDMRLRLRDPAMASAVEELRF-- 281
QY 301 LRIGT---VPEARODLEYCGASIKKGENVFLILSALRDGVTSRDPVFDVRPDTGASL 356
DB 282 IPLGSGSFRAVATPEVLECGRIOPGVTVPFTVAAMWDPVFAEPGLRIDRSPNPHV 341
QY 357 AYGRGPHVCPGVSLARLEAIVAGTIFPRFPMKL 391
DB 342 AFGHGVHRCGLAQRLARLEIQVALGVLLRRLPRRL 376

RESULT 15
Q8KNK8 PRELIMINARY; PRT; 394 AA.
AC Q8KNK8;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CALE10.
GN CALE10.
OS Micromonospora echinospora (Micromonospora purpurea).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micromonosporaceae; Micromonosporaceae; Micromonospora.
OX NCBI_TaxID=1877;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 15839;
RA Ahlert J., Shepard E., Lomovskaya N., Zazopoulos E., Staffa A.,
BA Bachmann B.O., Huang K., Feinstein L., Cziorny A., Whitman R.E.,
RA Farnet C.M., Thorson J.S.;
RT "The calicheaactin gene cluster and its iterative type I PKS";
RL Science 0:0-0(2002).
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
DR EMBL; AF497482; AAM94800.1; -.
DR InterPro: IPR001128; Cytochrome_P450.
DR Pfam; PF00067; p450; 1.
DR PRINTS; P800385; P450.

[illegible]


```

QY 130 VQOLDARSGQEPDVVRDYABEIPMKRAISALKVPAECDEKRRRGASATRALGVGVLP 169
DB 128 TDLLDRLGRSSVVDLIDEPAPLPITVISELWGVDSRRDRDPR---TNVLVDGSP 183
QY 190 QVDEETK-TLVASVTEGLALHVDLERRRNPENLVLTMLQAEADGSRSTKELVALV 248
DB 184 EAOQASVAMVEVLTSLIA-----KKRTGGDDLLTRLEAVEDGRLSPGELIANV 235
QY 249 GAIIAAGDTTTLTILAFVAVLNILRSPEALEVKAEPGLKRNALDEVLRFDNIIRICTVRP 308
DB 236 FLTLVAGHETTVNLINGCVSLILGNPDOLALNRDPSLLFGAIEETLRYESPVANGTFPH 295
QY 309 AAGDLEFCASIKKGMVFLILPSALRDGVFSRPDVPDRDGTASLAVGREGPHVCPGV 368
DB 296 TAEAVFPGDVVPEGLVWALCAANRDERFEDPDRFDTRETGTGHVAFGHGHCVA 355
QY 369 STARLAEIAGVTIFRRPEPKLETP---VFGYHAPFNIIESLNV 411
DB 356 ALARLEAQIAVGRVLRFPRDLRMAASPDRLRMRFVLMGLKLPV 401

RESULT 2
BIOI_BACSU STANDARD; PRT; 395 AA.
AC P53554;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Biotin biosynthesis; cytochrome P450-like enzyme (EC 1.14.-.-).
GN BIOI OR CYP107H.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96312354; PubMed=8763940;
RA Bower S., Perkins J.B., Yocum R.R., Howitt C.L., Rahaim P.,
RT "Cloning, sequencing, and characterization of the Bacillus subtilis
RT biotin biosynthetic operon."
RT J. Bacteriol. 178:4122-4130(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RA "sequencing and functional annotation of the Bacillus subtilis genes
RT in the 200 kb trnB-dnaB region."
RT Microbiology 143:3431-3441(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bettero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruneel C.V., Caldwell B., Capuano N.J., Daniel N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.D., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Ertan K.D., Errington J., Fabret C., Ferrati E., Follmer D.,
RA Fritsch C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaeser P., Goffeau A., Golligly E.J., Grandi G.,
RA Gaisbpi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasehara Y., Klaer-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Portollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,

```

```

RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seir S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toseato V., Uchiyama S., Vandenbol M., Vanlier F., Vassartoli A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,
RA Winere P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1. Similarity: Belongs to the cytochrome P450 family.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC CC
CC EMBL; U51868; AAB17462.1; -
CC DR EMBL; AF008220; AAC00266.1; -
CC DR EMBL; Z99119; CAB14997.1; -
CC DR PIR; G69594; G69594.
CC DR HSSP; Q55080; 1107.
CC DR Subtilist; BG11528; biot.
CC DR Interpro; IPR001128; Cytochrome_P450.
CC DR Pfam; PF00067; P450; 1.
CC DR PRINTS; PR00385; P450.
CC DR PROSITE; PS00086; CYTOCHROME_P450; 1.
CC KM Biotin biosynthesis; Oxidoreductase; Monooxygenase;
CC KW Electron transport; Heme; Complete proteome.
CC FT METAL 345 345 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
CC SQ SEQUENCE 395 AA; 44865 MW; E4AC3AE2637ACE1A CRC64;

Query Match 24.7%; Score 525.5; DB 1; Length 395;
Best Local Similarity 31.9%; Pred. No. 6.5e-26;
Matches 129; Conservative 98; Mismatches 137; Indels 41; Gaps 12;

QY 24 YAEPPPAERLREATPIFYWDEGR-----SWVLTRYHDSAVFDERFAVSREMESSA 78
DB 12 FLKNPFSFYDPLRAVPIY---KGSFLKYPGVVYGVETETALILKDAKFKVATPLPESST 68
QY 79 EYSSAIPESLDMKKYVLCGLPREDHARVRYKYNPSTSAIDLLAEIORTDQLDARS 138
DB 69 KYQ---DLSHVQNMMLFQONPDHRLKRTLASGAFTPTTSYQYIETVHLLDQVO 124
QY 139 GQEPDVVRDYABEIPMKRAISALKVPAECDEKRRFGSATRALGVGLVPOVDEETKTL 198
DB 125 GKKEVEVSDPAFLPASPITANIGVEEDRQLKEMAS-----LIQITD-PTRRR 175
QY 199 VASVTEG-----LALHVDLERRRNPENLVLTMLQAEADGSRSTKELVALVGI 251
DB 176 KA-LTEGINAVQAMVYFKELIOKRKRP-QQDMISMLKGR-EKKLTREEBAASCTIIL 232
QY 252 IAAAGDTTTLTILAFVAVLNILRSPEALEVKAEPGLKRNALDEVLRFDNIIRICTVRP 311
DB 233 AIAAGHETTVNLISNSVLCIQLHEQOLKREPNDLGTVERGLRESPTQM-TARVAS 291
QY 312 DLEYGASIKKGMVFLILPSALRDGVFSRPDVPDRDGTASLAVGREGPHVCPGV 371
DB 292 DIDICVTRRGQVVLVGLAANRDERFEDPDRFDTRETGTGHVAFGHGHCVA 355
QY 372 RLEAEIAGVTIFRRPEPKLETP---VFGYHAPFNIIESLNV 411
DB 352 RLEAQIAVGRVLRFPRDLRMAASPDRLRMRFVLMGLKLPV 392

RESULT 3
YUIB_BACSU STANDARD; PRT; 396 AA.

```

AC 034374;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative cytochrome P450 yj1B (EC 1.14.-.-).
 GN yj1B
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Rivalta C., Soldo B., Lazarevic V., Joris B., Manuel C., Karamata D.,
 RT "A 35.7 kb DNA fragment from *Bacillus subtilis* subfills chromosome containing a
 RT putative 12.3 kb operon involved in hexuronate catabolism and a
 RT perfect catabolite-responsive element."
 RT Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertier M.G., Besieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funas S., Galizzi A., Galleron N.,
 RA Chim S.Y., Glaier P., Goffeau A., Goldschly E.J., Grandi G.,
 RA Guisephi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Landus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Maeda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portwillik S., Prescott A.M.,
 RA Presacian E., Pulic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Satale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpsira P., Tognoni A.,
 RA Toso V., Uchiyama S., Vandenbol M., Vannier F., Vassaretti A.,
 RA Viati A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*."
 RT Nature 390:249-256 (1997).
 RL -1- SIMILARITY: Belongs to the cytochrome P450 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AF015825; AAC46317.1; -
 CC EMBL, Z99110; CAB13078.1; -
 CC PIR, B69851; B69851.
 CC HSP, Q85060; 1109.
 CC Subtilist; BG1195; yj1B.
 CC InterPro; IPR001128; Cytochrome_P450.
 CC Pfam; PF00067; P450; 2.
 CC PRINTS; PRO0365; P450.
 CC PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
 KM Complete proteome.

FT METAL 349 349 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 SQ SEQUENCE 396 AA; 44990 MW; 9A89CF12613DBCFC CRC64;
 Query Match 24.4%; Score 518.5; DB 1; Length 396;
 Best Local Similarity 32.6%; Pred. No. 1.8e-25;
 Matches 124; Conservative 72; Mismatches 155; Indels 29; Gaps 6;
 QY 25 AEDPPATERLEATPIFYWDEGRSVLTRYHDSAVFRDERFAVSREMESSAEYSSAI 84
 DB 23 AHFPFWESMKDAPVSGFDEENQVSVFLVDVKVVGDKL-----FSSGCM 70
 QY 85 PELSDMKYGLFGLPEPDHARVKLVNPSFTSRALDLRAEIQRYVDLLDRSGOEED 144
 DB 71 POTSISGINSINMDPPKHTKRSVYNKAFPRWKOWEPRIOETIDELIQFGOSSEPD 130
 QY 145 VVRDVAEGIPMAISALIKVPAECCEKPRFSGATAPALGVGLVQVDEETLVA---- 200
 DB 131 LVHDSYPLPVVISLGLVPSAHMEQFKASD-----LVSTPDKSEAEKAFLEED 185
 QY 201 SVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRSTKELVALVGAIIAGTDTTI 260
 DB 166 KCEELAAFPAGIIIEKKNKPEQDIIISLVEAEETGKLSGSELLPFTLLVAGNETTT 245
 QY 261 YIIAPAVNLRSPPALDLVKAEPGLKRNALDEVURF--DNILRIGTVRFAROLEYCG 317
 DB 246 NLISNAMYSLIETPPGVYELSHPELMPQAVEALRFAPAPVLR-----RIKRPDEICG 301
 QY 318 ASIKKGEVFLIIPALADGTFSRDPVDPVDRDGSALVGRGHVCGVSLARLENI 377
 DB 302 HIKKGDVLAFAVSAANDDEAKFDRPHMFDIRRHPHIFAGHIFGLAPLARLEANI 361
 QY 378 AVGITFRFRPEMK-LKETPV 396
 DB 362 ALTSLISAFPHMECVSTIPI 381

RESULT 4
 ID CPXY_BACSU STANDARD; PRT; 410 AA.
 AC 008469;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cytochrome P450 (EC 1.14.-.-).
 GN CYP4 OR CYP107J1.
 OS *Bacillus subtilis*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1.
 RX MEDLINE=97431495; PubMed=9287000;
 RA Bellamy B.R., Gustafsson M.C.U., Sonenshein A.L., von Wechenfeldt C.;
 RT "An lrp-like gene of *Bacillus subtilis* involved in branched-chain
 RT amino acid transport."
 RT J. Bacteriol. 179:5448-5457 (1997).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=97453479; PubMed=9308178;
 RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
 RA Dusterhoft A., Ehrlich S.D.;
 RT "Sequence of the *Bacillus subtilis* genome region in the vicinity of
 RT the lrp operon reveals two new extracytoplasmic function RNA
 RT polymerase sigma factors Sigv and Sigz."
 RT Microbiology 143:2939-2943 (1997).
 RL [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertier M.G., Besieres P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,

[illegible]

```

Db      255 NLIINMYALMCHHDQJEXKROQDPDLNMSAIEBALRHSHSVELTTIRWTAEPLILNGQEI 314
Qy      321 KKGEMVLLIPSAIRDGVTSRPDVEDRDGTASLAYGEGVHCPSIARLEAEIANG 380
Db      315 KKKVITISLASANRDEKIFPNADIFIEKRNRRHIAFGHGNFICGACQARLEAKIAIS 374
Qy      381 TIFRRPEPMKX 392
Db      375 TLRRCPNIOLK 386

RESULT 5
ID    CFX_STRGO      STANDARD;      PRT;      402 AA.
1D    STRAIN=ATCC 11796;
AC    P18327;
DT    01-NOV-1990 (Rel. 16, Created)
DT    01-NOV-1990 (Rel. 16, Last sequence update)
DT    28-FEB-2003 (Rel. 41, Last annotation update)
DE    Cytochrome P450-SU2 (EC 1.14.-.-) (P450-CYB1) (CYP105B1).
OS    CYP105B1 OR SUBC.
OS    Streptomyces griseolus.
OC    Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC    Streptomycinae; Streptomycetaceae; Streptomyces.
OX    NCBI_TaxId=1809;
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 1-32.
RC      STRAIN=ATCC 11796;
RX      MEDLINE=90264332; PubMed=2345149;
RA      Omer C.A., Lanstra R., Little P.J., Dean C., Tepperman J.M.,
RA      Leto K.J., Ramesser J.A., O'Keefe D.P.;
RT      "Genes for two herbicide-inducible cytochromes P-450 from
RT      Streptomyces griseolus.";
RL      J. Bacteriol. 172:3335-3345(1990).
CC      -1- FUNCTION: METABOLISM OF A NUMBER OF SULFONYLUREA HERBICIDES.
CC      -1- INDUCTION: By herbicides.
CC      -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M32239; AAA26825.1; -.
DR      PIR; B35401; B35401.
DR      HSSP; P23295; ICNM.
DR      InterPro; IPR001128; Cytochrome_P450.
DR      Pfam; PF00067; P450.1.
DR      PRINTS; PR00385; P450.
DR      PROSITE; PS00086; CYTOCHROME_P450.1.
KW      Oxidoreductase; Monooxygenase; Electron transport; Heme.
FT      INIT_MET 0
FT      METAL 351 351 IROn (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ      SEQUENCE 402 AA; 44278 MW; E3B67F6F72C6D9D CRC64;

Query Match 21.6%; Score 459; DB 1; Length 402;
Match local similarity 31.7%; Pred. No. 9,3e-22;
Matches 127; Conservative 73; Mismatches 157; Indels 44; Gaps 12;

Qy      21 APGYAEDPFAIERLREATPI--FYWDEGRSWYLRHYDVSAVFRDERFAVSREWESS 77
Db      19 APGCPDPADPVDTEARTEPRTATRLWDGSSCMLVTRHGVRAVLGDPFR-----SA 70
Qy      78 AEYSSAIPELSDMKKYGFLGIP-----EDHARVAKLYNPSTSAIDLLRAEIORTV 130
Db      71 DAHRTGPEFLTAGGR-SIGTNPFLRMDDPEHARLRMLTADITVKKVEMRREYQLA 129
Qy      131 DQILD-ARSGOEPDVVDVYEGIPMRISAALKVPAECCEKFRPFSGSATARAAGVLVP 189
Db      130 DDLVDKMTGTGTSADVLTEFAFLPSPSLVITCLLGVPE-DHAF--FQERSVLLTASTP 186

```

```

QY 190 QVDEETKTLVASVTEGIALHLDVLDERRRNPENLDVLTMLQADGSRISTKELVALVG 249
DB 187 ---EEVRAQDELLEYLARL---ARTKRERPDAAISRLVARGELDDTOATWGRLLLV- 239
QY 250 AIIAGTDTTIIYIAAVNLIRSPALVELVAKEPGLMRNALDEVLPFNIIIRIGTVRA 309
DB 240 ---AGHETTANNTALSTVLRLNPNQRLRLRAEPALVGAABELRLRYLTIVNGVPRIA 295
QY 310 RODLEYCGASIKKGEVWFLIPGALRDGVFSRPDVFDRDGTGASLAVGRGPHVCGVS 369
DB 296 TEDVLIGRTIAGEVLCMISSANDAEVFGCDLDDVARDRRRHVAFGVHQCLGQP 355
QY 370 LARLEAIAVGTIFRRFPPEKLT---KETP-----VFQYH 400
DB 356 LARVELQIAIETLRLRLPDLRLAVPHEIIPFGDMALYGVH 396

RESULT 6
C123 MYCTU STANDARD; PRT; 402 AA.
AC P77902;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 123 (EC 1.14.-.-).
GN CYP123 OR RV0766C OR MT0790 OR MTCY369.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OK NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fieischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterlind T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to the cytochrome P450 family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 280226; CAB02396.1; -
CC EMBL; AE006970; AAK45032.1; -
CC PIR; A70707; A70707.
CC HSSP; 000441; 10XA.
CC TIGR; MT0790; -.
CC Tuberculest; RV0766c; -.

```

```

DR InterPro; IP001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Complete proteome.
FT METAL 350 350 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 402 AA; 45421 MW; 76AD843019361798 CRC64;

Query Match 21.2%; Score 450; DB 1; Length 402;
Best Local Similarity 32.4%; Pred. No. 3,4e-21;
Matches 128; Conservative 64; Mismatches 179; Indels 24; Gaps 9;

QY 13 PADEFKPAAGYAEPPFAIERLRATPIFYWDGSRSWLTTRYHDSAVAFRD-----ERF 67
DB 8 PELVLDPIYDHDHPYRYRLRDEAPLYRNEENFMAVSHHHVLDQGFRTSLSNAY 67
QY 68 AVSREWESSAAYSAIPELSPMKYGLFGLEPDEHARVYKLVNPSFSAIDLLRAEIQ 127
DB 68 GVS---LPPSSRTSAYRVMS-----MLAMDPPALHMRKTLVSKGFTFRRIRELPQVL 118
QY 128 RTVDQLDARSQGEEDVAVROVYAEGLPMRAISALLKVPACDEKRRFSSATR-RALGVG 186
DB 119 ELARJHLDSALQTESFDFAEFAGKLPMDVISELIGVPTDRAIRALADAVLHRDGYA 178
QY 187 LVPQVDEETKTLVASVTEGIALHLDVLDERRRNPENLDVLTMLQADGSRISTKELVA 246
DB 179 DVP-----PPMAAASL-ELMRYADLIAEFRRRP-ANNUTSLALAEALDGDRLSDEIMA 231
QY 247 LVGAIIAGTDTTIIYIAVNLIRSPALVELVAKEPGLMRNALDEVLPFNIIIRIGTV 306
DB 232 FLFLWVIAGNETTTLKLANAVYMAAHPQQLARVPADSHRIPMWVEETLRYDTSSQI-LA 290
QY 307 RPARDLEYCGASIKKGEVWFLIPGALRDGVFSRPDVFDRDGTGASL-AVYCGPHVC 365
DB 291 RTVAHDLTLYDTTIEGEVLLLPSSANRDDRVPDDPDYRIGRIGKLVFSGGAFHC 350
QY 366 PGVSLARLEAIAVGTIFRRFPPEKLTETPVGVH 400
DB 351 LGAHLARLEAIAVGTIFRRIRNRYEVDDNVGVH 395

RESULT 7
C140 MYCTU STANDARD; PRT; 438 AA.
AC 008454;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 140 (EC 1.14.-.-).
GN CYP140 OR RV1880C OR MT1929 OR MTCY180.38.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OK NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gass S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultson J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```

RA Frieschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwin M.L., Hatt D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Belcher A., Ueberback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 CC Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z97193; CAB10066.1; -
 DR EMBL; AE007049; AAK46201.1; ALT_INIT.
 DR PIR; E70515; E70515.
 DR HSSP; Q00441; 10XA.
 DR TIGR; M11929; -.
 DR Tuberculist; Rv1880c; -.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KM Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
 KW Complete proteome.
 FT METAL 381
 SQ SEQUENCE 438 AA; 48871 MW; 0463B3EFCF58744 CRC64; (BY SIMILARITY).
 Query Match 20.5%; Score 434.5; DB 1; Length 438;
 Best Local Similarity 29.6%; Pred. No. 3.Se-20;
 Matches 126; Conservative 77; Mismatches 168; Indels 55; Gaps 13;
 QY 22 PGYADDFPAIERLREATPIFYWDEGRSWLTRYHDS-AVFRDERPAVSREWESSAEY 80
 DB 35 PAVATDPVPEYDEKRSGLAV---RNRANLYTDRLAHLNLSDDRRV-----VSF 83
 QY 81 SSAIP-----ELSDMKKYGFLGPPEDHARVKLVNPFSTSAIDLRAEIQ 127
 DB 84 GENLPPLRLWERTRGDQLHPLREPSLAVPEPDHTRRYKTVSAVFTSRVAVSLRDVLE 143
 QY 128 RTVDQLDARSQGEF-DVVRDYAEGIPMRASLALKVPAECDEKFRFGSATARALGVG 186
 DB 144 QTAIINLDRFAEQGVIVDVGRCYCSQLPIVVISLIGVPHDRPVLEFGLAAPSIDIG 203
 QY 187 LVQVDEETKTLVAVTEGL---ALHADVLDERRRNPLENDVLTMLQAEADG---SRL 239
 DB 204 I-----PMROYLVQGGIRGPDWCLEGLHQQRHNP-GDILMSQLIQIABSGDNNTQL 255
 QY 240 STYELVALGAIIAAGTDTTYYLIAFAVLNLLSPALVELVKAEPGLMRNALDEVIRFDN 299
 DB 256 DETELRAIAGLVAVAGFETTVNLGNGIRMLDTPBEHLATLRQHPMLPMTVEIIRLDS 315
 QY 300 IIRIGTVRPAODLEYGCAIYKKGEMVFLIPSLRGCTVPSRPDVDRV-DTGASLAY 358
 DB 316 PVQL-TARVACRDEVAVGRIKGEVVVYILAAANRPDAVFPDHRDIRPAGRHIAF 374
 QY 359 GRPAPHVPGVSLARLEAIVAGTIFRRFPEMKL-----KETPVFGYHAPFARNIESLNI 412
 DB 375 STGRHFLGALALRAEVEGLRTFFDFPVRAGAASRDTRV-----LRGMSITLPTV 428
 QY 413 LKPSKA 418
 DB 429 LGPARS 434

RESULT 8
 CPXM_BACSU STANDARD; PRT; 405 AA.

AC P27632;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cytochrome P450 109 (EC 1.14.-.-) (ORF405).
 GN CYP109.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 ON NCBI_TaxId=1423;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M23;
 RC MEDLINE=91192601; PubMed=1849493;
 RA Ann K.S., Wake R.G.;
 RT "Variations and coding features of the sequence spanning the
 RT replication terminus of Bacillus subtilis 168 and M23 chromosomes";
 RT Gene 98:107-112(1991).
 CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M24523; AAA22720.1; -
 DR HSSP; O5080; 1107.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KM Oxidoreductase; Monooxygenase; Electron transport; Heme.
 FT METAL 351 351 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 SQ SEQUENCE 405 AA; 45845 MW; 1331D5BEA74E3C04 CRC64;
 Query Match 19.8%; Score 420; DB 1; Length 405;
 Best Local Similarity 29.0%; Pred. No. 2.Se-19;
 Matches 121; Conservative 72; Mismatches 182; Indels 42; Gaps 9;
 QY 1 MTOEANOSETPKAPDFKFPAGYAD---PFAIERLREATPIFYWDEGRSWLTRYH 56
 DB 1 MTNQTARSSKKEKYANLIMEELHSKDLRFPPIYDKLRRESPVYDRLRCMDVFXKD 60
 QY 57 DVSATFRDERFAVSREWESSAEYSALPELSDMKKYGFLGPPEDHARVKLVNPFST 116
 DB 61 DVQFVLKNPKLPSKRGIQTES-----ILTMDPPGHTLRALVSPAFAP 104
 QY 117 RAIDLRAEIQRTVDQLDARSQGEFVDVVRDYAEGIPMRASLALKVPAECDEKRRRG 176
 DB 105 KAVKQLETRIKQVTAFLQEAROKSTIIDIEFAGPLPITIAEMGAIERHLLIKTYS 164
 QY 177 SATARALGVGLVQVDEETKTLVAVTEGLALP---HDVLDERRRNPLENDVLTMLQ 232
 DB 165 D---VLVAGAKDSSKAVADVHNRDGHAFSLDFRIILSKRAPEKE-DLMTLLDRA 219
 QY 223 EADGSLSTKELVALGAIIAAGTDTTYYLIAFAVLNLLSPALVELVKAEPGLMRNALD 292
 DB 220 EIDGEVLTBEQILGFCITLLVAGNETTVNLINAAVYLTEDSVQOVQVQONTDNVANVIE 279
 QY 293 EVLR-FDNLIRIGTVRPAODLEYGCAIYKKGEMVFLIPSLRGCTVPSRPDVDRV 351
 DB 280 ETLRYVSPVQAIG-RVATEDTELGGVFYIKGSSVSIWASANRDEKCKCDPCFKIDRP 337
 QY 352 TGASLAYGRPHVPGVSLARLEAIVAGTIF-----RPPPEMKLKEP---VFG 398
 DB 338 SYPHLSFGFHICLGLAPLARLEANIALSSLSMSACIKRAHDELTLEAIPEFPVVG 394

Proc. Natl. Acad. Sci. U.S.A. 97:3050-3055 (2000).

[5] X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).

RP MEDLINE=21363699; PubMed=11469860;

RA Cupp-Vickery J., Garcia C., Hofacre A., McGee-Estrada K.,

RT "ketocanazole-induced conformational changes in the active site of

cytochrome P450eryf."

RT J. Mol. Biol. 311:101-110(2001).

CC -1- FUNCTION: CONVERSION OF 6-DEOXYERYTHRONOLIDE B (6-DEB) TO

ERYTHRONOLIDE B (EB) BY THE INSERTION OF AN OXYGEN AT THE 6S

POSITION OF 6-DEB.

CC -1- PATHWAY: Erythromycin A biosynthesis.

CC -1- MISCELLANEOUS: A MUTATED FORM OF THIS ENZYME LEADS TO THE

FORMATION OF DEOXYERYTHROMYCIN A, AN ANTIBIOTIC WITH IMPROVED

PHARMACOLOGICAL PROPERTIES.

CC -1- SIMILARITY: Belongs to the cytochrome P450 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X60379; CAA42927.1; -

DR EMBL: M54983; AAA26496.1; -

DR PIR: S18531; S18531.

DR PDB: IOXA; 07-DEC-95.

DR PDB: IEUY; 02-APR-00.

DR PDB: LEUP; 28-APR-00.

DR PDB: LIJN; 17-OCT-01.

DR PDB: LIJO; 17-OCT-01.

DR PDB: LIJP; 17-OCT-01.

DR InterPro: IPR001128; Cytochrome_P450.

DR Pfam: PF00067; P450_1.

DR PROSITE: PS00086; CYTOCHROME_P450; 1.

KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;

KW Antibiotic biosynthesis; 3D-structure

FT METAL 351 351 IRON (HEME AXIAL LIGAND).

FT CONFLICT 263 263 Q -> ODQ (IN REF. 1).

FT CONFLICT 302 302 G -> R (IN REF. 1).

FT STRAND 6 6

FT TURN 7 8

FT HELIX 10 12

FT TURN 13 13

FT HELIX 16 26

FT STRAND 29 34

FT TURN 35 36

FT STRAND 37 42

FT HELIX 45 53

FT TURN 55 56

FT STRAND 57 58

FT TURN 61 62

FT TURN 66 67

FT HELIX 73 75

FT TURN 76 77

FT HELIX 80 86

FT HELIX 90 92

FT TURN 95 96

FT HELIX 97 106

FT TURN 107 108

FT HELIX 111 115

FT TURN 116 117

FT HELIX 118 132

FT STRAND 137 139

FT HELIX 140 143

FT TURN 144 146

FT HELIX 147 157

FT TURN 158 158

FT TURN 161 165

FT HELIX 166 175

FT HELIX 179 181

FT HELIX 182 205

FT HELIX 211 217

FT STRAND 219 220

FT TURN 221 223

FT STRAND 224 225

FT HELIX 228 229

FT HELIX 261 269

FT HELIX 271 273

FT HELIX 274 284

FT STRAND 291 295

FT STRAND 299 301

FT TURN 302 303

FT STRAND 304 306

FT TURN 308 309

FT STRAND 311 314

FT HELIX 316 319

FT TURN 320 321

FT TURN 323 325

FT TURN 329 330

FT TURN 334 335

FT TURN 339 340

FT TURN 343 344

FT TURN 348 349

FT TURN 352 353

FT HELIX 354 371

FT TURN 373 374

FT STRAND 376 377

FT HELIX 381 383

FT STRAND 386 386

FT STRAND 395 395

FT STRAND 399 401

SO SEQUENCE 404 AA; 45039 MW; 257ABEFC2D88B3FE CRC64;

Query Match 19.4%; Score 411; DB 1; Length 404;

Best Local Similarity 30.7%; Pred. No. 9, 2e-19;

Matches 119; Conservative 71; Mismatches 149; Indels 48; Gaps 14;

QY 24 VAEDEPPAERLEREATPI-----FYWDEGRSWLTRHYDVSAYFRDERFAVSREEMESSAE 79

DB 20 YAE-----LRETPVPVPRFLEGD--AWLVGTGDEAKALSDRLSSDCKKYPGYE 69

QY 80 YS-----SAIPELSDMKKYLFGI--PPEDHARVKLVNPSTSPRAIDLRAEIQTVDQ 132

DB 70 VEEPAVLGPPE--DVANYPATNMGTSDPPTHRLRLKLVSGEFTVVRVEMRRVEQITAE 127

QY 133 LIDARSGGEFDPVRYAGIIMRAISALTKYPAEDEKER--RFGSATRAALGVGLVPGV 191

DB 128 LIDVEGDSGVVDIVDFAPHLPLKYICELGV---DEKYRGDFGWSSEILVMD--PER 181

QY 192 DEETKTIVASVTEGLALHDVID--ERRNPLENVDLTWLLQA--EADGSRLSTKEIYALV 248

DB 182 AERQGAAREV-----VNFIDIVERRRTEPGDDLSALIVQDDDGRLADELTSTIA 235

QY 249 GAIIAAGDTTYLLIFAVLNLRSPEALELYKAEGKLRNALDEVLRFDNLRIG---- 304

DB 236 LVLLAGFASVSLIGITGLLTLPDQALVRRPSPALPNVBEILRY-----IAPPET 290

QY 305 TVRFARQDLEYGASIKKGEWFLIPSLRPDGTVSRRDPVRRDGTGASLAVGRGPV 364

DB 291 TTRFAAEVEIGVAVLPQYSTLVANGAANRDPKPPDHRPDVTRDTGHSFGGIGIF 350

QY 365 CPGVSLARLEAEIYAVGTIFRRFPEMKL 391

DB 351 CMGRPLAKLEGEVALALFGRRPALSL 377

RESULT 11

ID NOR2 CYLTO STANDARD; PRT; 408 AA.

AC Q12559;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)


```

QY 138 SGOEEFVVRADYVEGVPMAISALTKVPECQEKRRRGSATARLGVGLVPOVDEETKT 137
Db 118 CEEGCECDVRDLAAPPVMAVIGDMLCVREQCDMLKRSDDLVYFLSSHSVGEFQITMD 177
QY 198 LVASVTEGALLHDVLDERRRNPLENDVITMLLOEADGSRISTKEVALVGAIIAAGTD 257
Db 178 AFAPAYND---FTPATTAARRADPTD-DLVSVSVSEVGEERISLDELMEYTLIIIGDGE 233
QY 258 TTAYTLAPAVNLNLSRSPALVELVKAPEGMRVALDEVRF-----DNIRIGTVAFARQDL 313
Db 234 TTRHTLSGGTEQLLRKRDMDLIQDRPSLLPQATIEMLRTWAPVKNMRVULTA-----DT 288
QY 314 EYCGASIKKGEWVFLLI PSALRDGVFSRPDVFVDRDGTGASLAYGRGPHVCPGVSIARL 373
Db 289 EFGTGTALCGEKKMILLFEESANFDEAVFCFEPEKFDVQRPNSHLAFGFTGHCILGNQIARL 348
QY 374 EAEIAYGTIFRRPEEKL-KETPFVGYHA--FRNISLANYLWPS 416
Db 349 ELSLMERVYKRLPDRULVAADSVLPLRPANVSVGSESPVAVFTDS 394

```

```

RESULT 13
CPXG_STRSQ      .
ID_CPXG_STRSQ   STANDARD;      PRT;      381 AA

```

DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome P450_105C1 (EC 1.14.-.-).
GN CYP105C1 OR CYP105C1.
OS Streptomyces sp.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
NCBI_TaxId=1931;

FN SEQUENCE FROM N.A.
 RP [1]
 RX MEDLINE=90299781; Pubmed=2361941;
 RA Horii M., Ishizaki T., Paik S.Y., Manome T., Murooka Y.,
 RT "An operon containing the genes for cholesterol oxidase and a
 cytochrome P-450-like protein from a *Streptomyces* sp.",
 J. Bacteriol. 172:3644-3653(1990).
 CC - SIMILARITY: Belongs to the cytochrome P450 family

CC This SMS-S-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL; M31939; AAA26718.1; -.
DR HSSP; 000441; IOXA.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme.
FT METAL. 330 330 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 361 AA; 41703 MW; 841B959C9DDEA99C CRC64;

Query Match 17.8%; Score 378; DB 1; Length 381;

Best Local Similarity 30.3%; Pred. No. 9.8e-17;
Matches 115; Conservative 60; Mismatches 148; Indels 56; Gaps 12;

28 PFPALIERLREATPI---FYWDEGRSMLTRYHDVSAVFRDERFAVSREEMESSAEYSIAI 84

Db 19 PRAEMQALRHKAPVTRTAFAADGRPGMLVTGYSARAVLSDSRF-TARGERE-----HPAV 72

85 PELSDMKKYGFLGPEDHARVRKLVNPSFTSRAIDLRAEIQRVVDQLDARSGQ-EEF 143

```
Db      | : : ||| : : : : : |||
73 PRAATL-----EDE-RCRRLIAGQFARMROITGRTERTVREHLDAMEHMSP 121
```

```

Oy      144  DVVRDVADEBEGIPKRAISALLKPYAEODEKFR-----FGSATRALGVGLVQVDEETX 156
      122  DLVHNFALVPBSVIAIELLGVPPRRHFQIDILRMGFGFRST-----EE-- 166
Db      197  TLVASTEGCALIHDVLYDE---RRRNLENDVLTMLIQEADSGSRSTELVALVGAT 252
      167  -----VTBEFVSLIGQOLRLVRLKTEP-GDDLISGLIAMD--PALTDELSIAIEALL 217
Oy      253  AAGDTTYTILIAFAVYLNILRSPEALEYKAEFGILRNALDEVLRFNDULIEIGTVRFAROD 312
Db      218  VAGHGTTAQLIALQALGFLLLENHDDQLALRAAPALTESAVBELIRHLSVHNHGPRALOD 277
Oy      313  LEYCGASIKKGEVNFLLIPSLARDGVTSRPVDVFRDRTGASLYAGRGVHCBSGLAR 372
Db      278  ADIEGTPPVAGEVVVVVSGAANRRDPAREERBDADVTRKEDTGILAFGHGMHQCIGROLAR 357
Oy      373  LEAEIAVGTIPRRPEMKL 391
Db      338  IEURVALTALIERFPHILKL 356

```

RESULT 14	
NORI_CYLTO	
ID_NORI_CYLTO	STANDARD;
	PRT; 403 AA

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Cytochrome P450 55A2 (EC 1.14.-.-) (Cytochrome P450NR1).
 GN CYP55A2
 OS Cylinodrocarpon fontinale (Cylinodrocarpon lichenicola)
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; mitosporic Hypocreales; Fuariaium
 NCBI_TaxID=42744;

FN SEQUENCE FROM N.A.
 RP
 RC STRAIN=IFO 30561;
 RX MEDLINE=97163854; PubMed=9010609;
 RA "Kudo T., Tomura D., Liu D.L., Dai X.O., Shoun H.,
 RT "Two isoforms of P450_{chl} of *Cylindrocapsa kontanense*: molecular
 RT cloning of the cDNAs and genes, expressions in the yeast, and thio-
 RT putative NAD(P)H-binding site.";

CC -I- SMILAIRITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial

CC entities requires a license agreement (see <http://www.lsd-bld.cn/announcement/>)
 CC or send an email to license@lsd-bld.cn.
 CC -----
 DR EMBL, D78511; BA011408.1; -.
 DR HSSP, P23295; IF24.
 DR InterPro, IPR01128; Cytochrome_P450.
 DR Pfam, PF00067; P450_2.
 DR PROSITE, PS00086; CYTOCHROME_P450; FALSE_NEG.
 DR Oxidoreductase; Monooxygenase; Heme.
 FT METAL 352 352 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT -----

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

	Query Match	Best Local Similarity
17.7%;	28.0%;	
Score 376.5; DB 1;	Pred. No. 1.3e-16;	
Length 403;		

Matches	116;	conservative	59;	mismatches	184;	indels	35;	gaps	10;
---------	------	--------------	-----	------------	------	--------	-----	------	-----

```
QY      13 PARDFKPAPGGAEDPFPALEKREAFIF---YWDEGRSWLIRIHVSASFKEURFAV 69
```

```

6 p$F---p$Q$R$G$M$P$P$A$F$A$K$U$R$A$I$D$P$V$K$V$K$E$D$G$S$L$A$W$E$V$I$K$I$K$O$V$I$F$V$A$I$D$E$R$L-- 60

```

QY 70 SREWESSAEYSSALPELSDMCKYGLFGLF-----PEDEHAKVKLVNFSISKALDEUN 123

3.

```

Db 61 -----SKVTRPGPELNLAGGKQAKAPTVNDADPHNNQKGVESLFLHEHYKIQ 114
Qy 124 AEIQRTVDQLDARS-----GQEPFVVRDYGEGIPRAISALLKVP-----AECDEKPR 173
Db 115 PYIKTVDDLLAMKKKGCANPVDLVKEFALPVPSYIIITLGGVFPNDLHLTQNNMR 174
Qy 174 RFGSATARLGLVLPQVDEEFTKLVASSTEGALLHDVLDERRRRNPLENDVLTMLQAE 233
Db 175 TNGSSTARAS-----AANQELLDVLAISVERLEEP-KDDLSIKLCTEQ 218
Qy 234 ADGSLSTKEIYALVGAIIAGTDTIYIAFAVNLARSPEALEYKPEGLMNADE 293
Db 219 VKPGIEKADAVQIAFLVLVAGNATVNNIAGVTLFQHPQLAQKXNPSLADQFVEE 278
Qy 234 VLRFDNLRIGTVRPARODLEYCGASIKKGEVFLIPSLBDGTVFSRPOVDFVRDPTG 353
Db 279 LCRVHTASALAIKRNKVDLEIGKHIKANEGLIANSQANDADQIFENPDEFNMRKRP 338
Qy 354 AS--LAYGRPHVPCGVSLARLEAIVGTIRRPPEM-----KLKETPVFG 398
Db 339 AEDPLGYGFPRIACIEHLAKELTTFVATLFKEFPDLINAVPEKINFTPLGG 392

```

RESULT 15

C128_MYCTU STANDARD; PRT; 489 AA.

```

AC Q59572;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cytochrome P450 128 (EC 1.14.-.-).
GN CYP128 OR RV2268C OR MT2330 OR MTCY339.42.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxId=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gae S., Barry C.E. III, Tekala F.,
RA Baddock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jorgensen K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -! SUBCELLULAR LOCATION: Integral membrane protein.
CC -! SIMILARITY: Belongs to the cytochrome P450 family.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

DR EMBL; 277163; CAB00967.1; -.
DR EMBL; AE007076; AAK46612.1; -.
DR PIR; H70729; H70729.
DR HSSD; 000441; 10XA.
DR TIGR; MT2330; -.
DR TIGR; RV2268C; -.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein; Oxidoreductase; Monooxygenase; Heme;
KW Transmembrane; Complete proteome.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 315 335 POTENTIAL.
FT TRANSMEM 383 403 POTENTIAL.
FT METRL 435 435 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ
SEQUENCE 489 AA; 5313 MW; 633F233CEFD3AD7A CRC64;

```

Query Match 17.6%; Score 373.5; DB 1; Length 489;
 Best Local Similarity 28.3%; Pred. No. 2.5e-16;
 Matches 115; Conservative 65; Mismatches 204; Indels 23; Gaps 6;

```

Qy 16 DFKPAPGYADPPPAIRLEATPIFYWDGSRVLTTRYHDVSAVFPDERPAVSREWE 75
Db 101 DFDPPDAIADPPVPHYRELLAGERVQYPRDVIYLSRYADVBEAARNHDTLSARGVT 160
Qy 76 SSAEYSSAIPPELSOMKKGGLGRLPEDHARVRLVNPFSRTAIDLAAEIQRTVDQLD 135
Db 161 FSRGMLPFLP-TSD-----PPATNRKQQLAPGMARGALETWPMVDQLARELVG 209
Qy 136 ARSGCEBPVDVRYAEGIPMRASALLKVPACDEKFPFGSATARALGVGLVPQVDEBT 195
Db 210 GLTQTPADVSTVAAPRPAITSVLCVDPDEAFCRLSNQARRITDVAL-----S 262
Qy 196 KTLVASVTGLA---LHDVLDERRRNPLENDVLTMLQAEADGSRISTEYALVGAII 251
Db 263 ASGLISTVQGFAGFRRRLALFTHRDNLRECTVIGKLTATHAEGRLSDDELFFFAVL 322
Qy 252 IAAGDPTIYIAFAVNLARSPEALEYKPEGLMNADEVFNDILNIGTRFARQ 311
Db 323 LVAGESTAHMIISTFLTLADYPOQLTLAQOPLIPSAIEHNRFPISPIQ-NICRTTRV 381
Qy 312 DLEYCGASIKKGEVFLIPSLBDGTVFSRPOVDFVRDPTGASLAYGRPHVPCGVSLA 371
Db 382 DYSVQANIPAGSLVLLAMGAAANDPROYEDPVPFRADRNRVGHILAFSGIHLCTGTOLA 441
Qy 372 RLEAEIAGTIRRPPEMKLKEPVGYPAPRNIESLNVILKPSKA 418
Db 442 RMEGQALIRELVANIDRIEVEVPPTWTJNANLRLGLTRVAVTPRVA 488

```

Search completed: October 2, 2003, 17:28:42
 Job time : 9.74239 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comogen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:33 ; Search time 14.5739 seconds
(without alignment)

2764.850 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124

Sequence: 1 MTOEQANQSEKPAFDKPKF.....HPAFNIESLNLKPSKAG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	531	25.0	405	1 B42606	cytochrome P450 CV
2	525.5	24.7	395	1 G68594	cytochrome P450 BI
3	518.5	24.4	396	1 B63851	cytochrome P450 YJ
4	502	22.6	410	1 E63611	cytochrome P450 CY
5	486	22.9	376	1 G63679	polyketide hydroxy
6	481.5	22.7	418	2 G83229	cytochrome P450 PA
7	459	21.6	403	2 B35401	cytochrome P450 10
8	451	21.2	415	2 AC2731	cytochrome P450 HY
9	451	21.2	415	2 B97512	probable cytochrome
10	450	21.2	402	1 A70707	cytochrome P450 RV
11	442.5	20.8	434	2 C87170	probable cytochrome
12	434.5	20.5	438	1 E70515	cytochrome P450 RV
13	430	20.2	411	1 J04003	cytochrome P450 -
14	429	20.2	417	1 S49051	cytochrome P450 CY
15	429	20.2	411	2 T35526	probable cytochrome
16	421.5	19.8	444	2 C83336	probable cytochrome
17	420	19.8	411	1 J05859	polyketide synthase
18	415.5	19.6	402	2 H82813	cytochrome P450-11
19	413	19.4	397	1 S51594	cytochrome P450 MY
20	411	19.4	404	2 T30231	cytochrome P450
21	406	19.1	410	2 T44767	cytochrome P450
22	400	18.8	406	2 S18531	cytochrome P450
23	386.5	18.2	408	2 P00007	cytochrome P450 ER
24	386.5	18.2	408	2 J05674	cytochrome P450 NO
25	386	18.2	398	1 H70807	cytochrome P450 RV
26	380	17.9	381	1 S15809	cytochrome P450 CY
27	373.5	17.6	489	1 H70729	cytochrome P450 RV
28	366	17.2	410	2 S39924	cytochrome P450me9
29	364.5	17.2	399	1 A55578	cytochrome P450 -

30	361.5	17.0	412	2 S24750	cytochrome P450 (a
31	361.5	17.0	413	2 T44587	cytochrome P450 HO
32	357.5	16.8	430	2 J04287	cytochrome P450 AC
33	356.5	16.8	400	2 B82817	cytochrome P-450 h
34	354.5	16.7	406	2 A35401	cytochrome P450 10
35	353.5	16.6	402	2 J05151	nitric-oxide reduct
36	353.5	16.6	404	1 J05150	nitric-oxide reduct
37	352.5	16.6	405	1 F69611	cytochrome P450 CY
38	350	16.5	401	1 I40208	cytochrome P450 BJ
39	344.5	16.2	400	1 H70921	cytochrome P450 RV
40	344.5	16.2	433	1 B70677	cytochrome P450 RV
41	342.5	16.1	428	1 A42971	cytochrome P450 ER
42	338.5	15.9	444	2 F87366	cytochrome P450 FA
43	332	15.6	410	1 O4BS6M	cytochrome P450 10
44	328.5	15.5	406	1 A48495	linalool 8-monooxy
45	328.5	15.5	467	2 T10879	y41c protein - Rhi

ALIGNMENTS

RESULT 1

B42606
cytochrome P450 CYP11B1 - Saccharopolyspora erythraea
N/Contains: oxidoreductase (EC 1.-.-.-)
C/Species: Saccharopolyspora erythraea
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Mar-2000
C/Accession: B42606

R/Author: J.F. Hutchinson, C.R.

J. Bacteriol. 174, 725-735, 1992

A/Title: Characterization of Saccharopolyspora erythraea cytochrome P-450 genes and enzy

A/Reference number: A42606; MUID:92121109; PMID:1733208

A/Accession: B42606

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: DNA; protein

A/Residues: 1-405 <AND>

A/Cross-references: GB:M83110; NID:G152682; PIDN:AAA26483.1; PID:G152684

A/Experimental source: NRR12338

A/Note: sequence extracted from NCBI backbone (NCBI:77484)

C/Genetics:

A/Gene: CYP107B1

C/Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology

C/Keywords: Chromoprotein; heme; iron; metalloprotein; oxidoreductase

F:238-374/Domain: cytochrome P450 homology <CYP>

F:352/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 25.0%; Score 531; DB 1; Length 405;

Best Local Similarity 32.5%; Pred. No. 6.8e-30;

Matches 132; Conservative 83; Mismatches 169; Indels 22; Gaps 6;

QY	16	DEKPAFGYADDPFAIRLRRE--ATPFYDEGRSWLTRYHDVSAVFRDER----	PAV 69
DB	8	DLLAFDDAFADRRNRVAKREBPVQRTYVGLDAMITREYDQALLPRLAKDKGR 67	
QY	70	SREWESSAEVSSAIPELSDMKYGLFGLPEDHARVRLVNPSTSAIDILRAEIORT 129	
DB	68	TQOIREKRLAARRRPGSPDLGRHMLNTDPPDRRLAKLVYKATARVEELRRRIQI 127	
QY	130	VDQILDARSQGEEDPVVDVAGIIPRAISALLKVPACDEKFRFGSATARALGVGLVP 189	
DB	128	TDDLLDRLRAGSEVDLIDFAFPLPIYISLMEGVDSRRDPRSM---TNVLVDGSP 183	
QY	190	QVDEETK-TLVASVTEGALLHDVDERRRPLENDVLTMLQAEADSRISTEVLV 248	
DB	184	EAQAQASVAMVEYLTETLA-----KKRTPEODLLTALLEAVDDSRISSEGLIMV 235	
QY	249	GAILAGDTTIVYLAFAVLNLRSPALIEVKAEPGLMRNALDEVLFNDILRIGTVRF 308	
DB	236	FLVLVAGHETTVNLIGNCVSLGNPDQALALRNDPSLPPAIEETLAVESPVANGTFRH 295	
QY	309	ARQDLEYGASIKKGEMVFLIIPSAIRDGVFSRPDVFVRRDVGASLAVGRPVCGV 368	
DB	296	TAEAVRFQDVVIPQELVWVALGAANRDEGFEDPDRFDITRETTGHAFAFGHIFCVGA 355	

C:Species: *Bacillus subtilis*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: E69611; T44774
R:Kunst, F.; Ogasaara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertone
C.: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Ch
A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gall
leech, J.; Harwood, C.R.; Henuit, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.
Richter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y. M.; Ogawa, K.; Ogikawa, A.; Oudsga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, E.; Rose, M.; Sadate, Y.; Sato, T.; Scanon
A:Authors: Schleich, S.; Schroeter, R.; Scalfone, F.; Segkuchi, J.; Sekowska, A.; S
akenchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumsattel, E.; Yoshikawa, E.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:960404033; PMID:9384377
A:Accession: E69611
A:Strutis: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <KUN>
A:Cross-references: GB:Z29117; GB:AL009126; NID:g2634966; PIDN:CB14615.1; PID:g2635119
A:Experimental source: strain 168
J:Belitsky, B.R.; Gustafsson, M.C.N.; Sonenshein, A.L.; von Wachenfeldt, C.
J. Bacteriol. 179, 5448-5457, 1997
A:Title: An lrp-like gene of *Bacillus subtilis* involved in branched-chain amino acid tre
A:Reference number: Z22837; MUID:97431495; PMID:9287000
A:Accession: T44774
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-410 <BEL>
A:Cross-references: EMBL:Y11043; NID:g1926275; PIDN:CAA71937.1; PID:g1926278
A:Experimental source: strain 1A1
C:Genetics:
A:Gene: *cypA*
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
P:245-381/Domain: cytochrome P450 homology <CYP>

```

Query Match Summary: 23.6%; Score 502; DB 1; Length 410;
Best Local Similarity 30.9%; Pred. No. 7,7e-28;
Matches 115; Conservative 88; Mismatches 159; Indels 10; Gaps 4;

QY 24 YAEDPFAIERLEREAT---PIFYWDEGRSWLTRYHDVASAVFERDERFAVSEEMESSAEY 80
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 22 FKDEAYEYFEKLRKSQALYPLSLGALGKGMILSRDDAILHLKCKKNGEYENFTAKEX 81
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 81 SSALPELSDMKKYGFLGFLRPEDHARVRLYNPSPFSRAIDILRLAEIQTTVDQLDANSQ 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 82 RPALCKEETLYTKMLNSDPDHNRLRLVOKAFPHRMILDLBCKIOHIAISLDDKVOYP 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 141 EEPDVDRVYAGGIPMRASALLKVPACCECKFRFRGSAATABALGVLGPOUDEETKTIVA 200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 142 KPMNLVDDYAPPLPIVISEMLGIFLEDRQKFRVWSQA---LTFSPDAPKRLQENDHILG 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 201 SVTEGALLHLDVDERRRNPLENDVLTMLLOAEADSGSLSTKEVALVGALIAAGTDTTI 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 199 EFVE---YLESILVKKRREP-AGDILSALIOAESGQSTSEBELYSMTMLIVAGHETTY 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 261 YLIAFAVLNLRSPALAEVLVKAEPGLMKNALDEVLRPNILRIGVRARADOLEVCASI 320
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 255 NLITNMTYALMCHHQLEKLRQOPDKMSALIEALRFPSPELTTIRRTAPPLIHGEI 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 321 KKGENVPLLIPSAIRDGTVFSRPDVFVDRBDTGASLARGPVPYCVSLARLEAEIAVG 380
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 315 KRKDVYIISLASARDEKIFPNADIFDIERKNNRIIARGHGNHFLGALRLAEKIAIS 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 381 TIFRRPFEMKX 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 375 TLRRCPNIOLK 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5
G69679
polypeptide hydroxylase (EC 1.-.-.-) pKes - *Bacillus subtilis*
C:Species: *Bacillus subtilis*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
J:Accession: G69679
R:Kunitz, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Bertozzi, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallardo, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hulo, M.F.; Koester, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kunitz, K.; Lapidus, A.; Lardinois, A.; Laubert, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Mauee, Y.M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli, R.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadle, Y.; Sato, T.; Scanlon, A.; Seccombe, A.; Seccombe, A.; Schleich, S.; Schroeder, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Serrano, A.; Serrano, A.; Serrano, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama, A.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, A.; Yoshikawa, H.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: G69679
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-176 <KUN>
A:Cross-references: GB:Z99113; GB:AL009126; NID:G26340930; PIDN:CB13607.1; PID:G2634107
A:Experimental source: strain 168
C:Genetics:
A:Gene: pKes
C:Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
F:240-376/Domain: cytochrome P450 homology <CYP>

Query Match	22.9%	Score 466	DB 1	Length 376
Best Local Similarity	31.6%	Pred. No. 9,2e-27		
Matches	120	Conservative	76	Mismatches 156; Indels 28; Gaps 7
QY	17	FKPPAAGYAEDPPPAIERLREATPIFYWDGR-----SWLTRYHYDVSAVRDERFA--	68	
DB	8	FHPHGKEHNHPFSLGFRFEPEPIHREFELKRGATYPAWLITRRYDDCAAFKDNRIITRD	67	
QY	69	---USREMESSAEVSSAIPELSDMKKYGGLGPPREDHARVRYKUNPSFTSRAIDLLRAE	125	
DB	68	VKNVNNQOIKMLNVSEIDIDFVSD---HMLAKTDPDTRLRLSLVHOAFTPTTIENLRGS	123	
QY	126	IORTVDOQLDARSGOEEFDVVRDYVAEGIPMRAISALTLVPAACEDEKFRFRFSATARAL--	183	
DB	124	IEQAIFQGLDMEKENKADIMKSFASPLPIVLSLWGI PREDRSQFOIWTMAAVDTSEIG	183	
QY	184	GVGLVPQVDEETKTLVASVTBGLALLHVDLDERRNPLENVDYITMLLOAEADGSLSTYE	243	
DB	184	NRELTNQALREFKQYIAK-----LIHD---RRIKP-KDPLISKLVAHEENGSKLSEKE	232	
QY	244	LVALVGATIAAGTDTTTLIAFAVLNLRSPEALETVAAEPLMNAALDEVIRFONITRI	303	
DB	223	LYSMLFLVAVAGLETTVLLSGSGLTALLQHKKECEKLKQOPEMITYABELLIRYSPVYM	292	
QY	304	GTVAFARODLEYCGASIKKGEVFPLLIPSLALDGTVFSRPDVFVDRDTGASLAVRGCPH	363	
DB	293	MANMWAIEDFYKGGHSIKRGDMIFIGISANRDPNFENPETLININRSPNRIISGFGIH	352	
QY	364	VCPGVSLARLAEAIYAGTIF	383	
DB	353	FCLGAPLARLEGGHIAFKAAF	372	

RESULT 6
G63229
cytochrome P450 PA331 (imported) - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 16-Feb-2001

C/Accession: G83229
 R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Badian, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam, J.; Loty, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A:/Reference number: AB2950; MUID:20437337; PMID:10984043
 A/Accession: G83229
 A:/Status: preliminary
 A:/Molecule type: DNA
 A:/Residues: 1,418 <STO>
 A:/Cross-references: GB:AE004755; GB:AE004091; NID:g9949456; PIDN:AA06719.1, GSPDB:GN001
 A:/Experimental source: strain PA01
 C/Genetics:
 A:/Gene: PA3331
 C:/Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology
 P:/333-369/Domain: cytochrome P450 homology <P45>

Query Match 22.7%; Score 481.5; DB 2; Length 418;
 Best Local Similarity 33.5%; Pred. No. 2,2e-26;
 Matches 116; Conservative 73; Mismatches 140; Indels 17; Gaps 5;
 QY 50 WVLTRHVSVAVRDERFPAVSREBESSAAY---SAIFELSDMKKYGFLGPEPDHAR 105
 DB 41 WVTTRYADAKVNLNHP--GVRDARQAALYAKKTSPPAGISGSHHMLNDPPDHTR 98
 QY 106 VRLVNPSTFRAIDLRARIORTVDQLARSGOEFVDVVRDVAEGIPMRATISALKVP 165
 DB 99 LRLVGRAPFPPOVERIQPIERTTEALDAMGRBADMADPAIPLTAVIFELGIP 158
 QY 166 -AECDEKFRFPGSATARALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRBNPLENDV 225
 DB 159 EAREHARQWEROAELLS-----PEEAQALADAVDVLRLV--LEAKRQPAD-DV 207
 QY 226 LTMLOAEADGSLSTKEVALYGAIIAAGTDTTIIAFAVNLRSPEALBYAKAPG 285
 DB 208 YSLGVQAADSGQLSEAEIVSMHLLMMSGFEETMMNIGALVTLLVNPQLALLRQPE 267
 QY 286 LHMNALDEVLPFNITLRIGVRFARODLEYCGASIKKGEWVFLIPSLADGTVFSPDV 345
 DB 268 LTRNAHEELVRHDSPVASMLRFTVEDVLDGVTIPAGEITLVSNLTANHDAERFDPPDR 327
 QY 346 FVVRDPTGASLAVGRPHVCPVSLARLEAIVAGTIFRRPEMKL 391
 DB 328 LDTLRNTDGLGFGFGVHYCVGASLARLEGRALIQRLARFPDQL 373

RESULT 7
 B35401
 cytochrome P450 105B1 - *Streptomyces griseolus*
 N:/Contains: oxidoreductase (EC 1.-.-.-)
 C:/Species: *Streptomyces griseolus*
 C:/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 04-Mar-2000
 C:/Accession: B35401
 R:/Owner, C.A.; Lénstra, R.; Little, P.J.; Dean, C.; Tepperman, J.M.; Leto, K.J.; Romesser, J.; Bacteriol. 172, 3335-3345, 1990
 A:/Title: Genes for two herbicide-inducible cytochromes P-450 from *Streptomyces griseolus*
 A:/Reference number: A35401; MUID:90264332; PMID:2345149
 A:/Accession: B35401
 A:/Molecule type: DNA
 A:/Residues: 1,403 <ONE>
 A:/Cross-references: GB:M32239; NID:g153480; PIDN:AAA26825.1; PID:g153481; GB:M36481
 C/Genetics:
 A:/Gene: CYP105B1
 C:/Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology
 C:/Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase;
 F:/238-374/Domain: cytochrome P450 homology <P45>
 F:/352/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.6%; Score 459; DB 2; Length 403;
 Best Local Similarity 31.7%; Pred. No. 8.1e-25;
 Matches 127; Conservative 73; Mismatches 157; Indels 44; Gaps 12;

QY 21 ARGYADPEPPAIERLREATPI---FYWDEGRSWLTRYHDSVAVRDERFPAVSREBESS 77
 DB 20 AGCCFPDPAVDYEAARTPTVRATIMDDSSCMLVTRHODVRAVLGDPRF-----SA 71
 QY 78 AERYSAIPELSDMKKYGLFGLP-----EDHARVKLVNSFTRALDILRAEIORV 130
 DB 72 DHRGTFFPLTYAGR-EIIGTNPFLRMDPPHARRRRLTADFTYKVKYAKNPEVQRLA 130
 QY 131 DOLL-D-ARSGOEFVDVVRDVAEGIPMRATISALKVPAECDEKFRFPGSATARALGVGLV 189
 DB 131 DDLVDMKTRTSADLVTEFALPLPSVLCLLGVPYE-DHF--FQESRVLTLRSPT 187
 QY 190 QVDEETKTLVASVTEGLALHDVLDERRBNPLENDVLTMLLOEADGSLSTKEVALYG 249
 DB 188 ---EEVRAAQDELLEIARL--ARTKRRPDPAISRLVARGELDDTOIATGRLTL- 240
 QY 250 AIIAAGTDTTIIAFAVNLRSPEALBYAKAPGEMNALDEVLPFNITLRIGVRA 309
 DB 241 ---AGHETTMNTALSTLVLRNPDQLARLEAPALVGAABELRYTLTVHNGVRLA 296
 QY 310 RODLEYCGASIKKGEWVFLIPSLARDGTVFSPRPVVRDPTGASLAVGRPHVCPGV 369
 DB 297 TEDVLIGFTIAGBGLCMISSANRDAVFPFGGDDLDVARARHVARFGVHQLGCP 356
 QY 370 LARLEAIVAGTIFRRPEMKL---KETP-----VFQYH 400
 DB 357 LARVELQIAIETLRLRLPLRLAVPHEIIPFGDMAIYGVH 397

RESULT 8
 AC2731
 cytochrome P450 hydroxylase Atu1256 [imported] - *Agrobacterium tumefaciens* (strain C58,
 C:/Species: *Agrobacterium tumefaciens*
 C:/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 C/Accession: AC2731
 R:/Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I.
 erage, G.; Gillel, W.; Grant, C.; Guenther, D.; Kutyavain, T.; Levy, R.; Li, M.; McCell
 /Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
 A:/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AC2731
 A:/Status: preliminary
 A:/Molecule type: DNA
 A:/Residues: 1,415 <KOR>
 A:/Cross-references: GB:AE008668; PIDN:AA042265.1; PID:g17739662; GSPDB:GN00186
 A:/Experimental source: strain C58 (Dupont)
 C/Genetics:
 A:/Gene: Atu1256
 A:/Map position: circular chromosome
 C:/Superfamily: *Bacillus* cytochrome P450 CYP106; cytochrome P450 homology
 C:/Keywords: heme; iron; metalloprotein
 F:/367/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.2%; Score 451; DB 2; Length 415;
 Best Local Similarity 29.1%; Pred. No. 3.1e-24;
 Matches 118; Conservative 82; Mismatches 175; Indels 30; Gaps 9;

QY 22 PGYADPPAIERLREATPIFYWDEGRSWLTRYHDSVAVRDERF-----AVSREB 75
 DB 24 PARVNDPFPVYALHAQCFYWEORQWFFCYDHSVSTLLDRRGRGROIHLVARSRE- 81
 QY 76 SSAEYSSAIPF-LSDMK-----KYLFGLPEPDHARVKLVNPSFTRALDILRAEIOR 128
 DB 82 -----IGPEPLEHKGFDLAEQHSLLLEPEHTRRLTLINRAVSHVDMKPEIEE 135
 QY 129 TVDQLDARSGOEFVDVVRDVAEGIPMRATISALKVPAECDEKFRFPGSATARALGVGLV 188
 DB 136 LARLLEAPBANGCETELSSYADIIPYMIARITGIPFEMGOLKMSHA---YGVWYMF 192
 QY 189 PQVDEETKTLVASVTEGLALHDVLDERRBNPLENDVLTMLLOEADGSLSTKEVALV 248

Db 193 KRTPEDELLADKAAQEFSDYVRVIAERRAEP-KODLLSHMHTSKGQYLTDDELVSTT 251
Qy 249 GAIIAAGDTTITYLIAFAVLNLRSPBEALVELYKABGRLMNLDEVLPRDNLIRIGTVRF 308
Db 252 IYVLNAGHEATVHOIGNSVRIITLESGLDPKTLFHDETATERTEVEETLRICAPVHI-FQKW 310
Qy 309 ARQDLEYCASIKKEMVFLIPSAIRDGTFSRDPVDRDGTASLAVGRCFVPCGV 368
Db 311 VLEPEIDGVQFKRDKVSLIAANLDPKFSDDLAFQPDNNEGANSFGAGIHFCIGA 370
Qy 369 SIARLEAEIAGVTIRRRPEMKLKETPVFG--YHPAFNIESLVN 411
Db 371 PLARLEMLALPLFLFKRLPGLKIAEPKVKDYH--FHGLERLDL 413

RESULT 9

B97512
Probable cytochrome P450 hydroxylase [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C:Accession: B97512
R:Goodner, B.; Hinkle, G.; Gattung, S.; Waller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollem, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: B97512
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-415 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87051.1; PID:g15156303; GSPDB:GN00169
C:Genetics:
A:Gene: AGR C 2319
A:Map position: circular chromosome
C:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: heme; iron; metalloprotein
F:367/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 21.2%; Score 451; DB 2; Length 415;
Best Local Similarity 29.1%; Pred. No. 3.1e-24;
Matches 118; Conservative 82; Mismatches 175; Indels 30; Gaps 9;
Qy 22 PGYADPPPAIRLEATPIFYWDGSRVNLTRYHDVAVFDERP-----AVGREWE 75
Db 24 PAFVNDPVPVYALAAQCTFYWEORQWFTCYDHVSTLDRFRGQILHVAARE-- 81
Qy 76 SSAEYSAIPE-LSDMK-----KYGFLGPLPEDHARVAKLVNPFSTRAIDLRAEIQ 128
Db 82 -----IGLPEPLEHVKHFDLAEOHSILEBEPHTRRLTLINRAFVSRHVKMPEIEE 135
Qy 129 TVDQDLARSGOEFVDVVAEGIPMRASIALKVPACDEKFRFGATRALAGVGLV 188
Db 136 LANRLIEAFANGETELISSVADILPVTMIAMIGIPPEMGQOLKWSHA---YVGWVF 192
Qy 189 PVDDEETKLVASVTEGLALLHDVLDERRRNPLENDVLTMLQAEADGSRSLSTKELVALV 248
Db 193 KRTPEDELLADKAAQEFSDYVRVIAERRAEP-KODLLSHMHTSKGQYLTDDELVSTT 251
Qy 249 GAIIAAGDTTITYLIAFAVLNLRSPBEALVELYKABGRLMNLDEVLPRDNLIRIGTVRF 308
Db 252 IYVLNAGHEATVHOIGNSVRIITLESGLDPKTLFHDETATERTEVEETLRICAPVHI-FQKW 310
Qy 309 ARQDLEYCASIKKEMVFLIPSAIRDGTFSRDPVDRDGTASLAVGRCFVPCGV 368
Db 311 VLEPEIDGVQFKRDKVSLIAANLDPKFSDDLAFQPDNNEGANSFGAGIHFCIGA 370
Qy 369 SIARLEAEIAGVTIRRRPEMKLKETPVFG--YHPAFNIESLVN 411
Db 371 PLARLEMLALPLFLFKRLPGLKIAEPKVKDYH--FHGLERLDL 413

RESULT 10

A70707
Cytochrome P450 Rv0766c - Mycobacterium tuberculosis (strain H37RV)
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Mycobacterium tuberculosis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A70707
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentile, S.; Hamlin, N.; Holroyd, S.
; Rajadaram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70707
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-402 <COL>
A:Cross-references: GB:Z80226; GB:AL123456; NID:g3261638; PIDN:CAB02396.1; PID:g1550644
A:Experimental source: strain H37RV
C:Genetics:
A:Superfamily: Bacillus cytochrome P450 CYP106; cytochrome P450 homology
C:Keywords: oxidoreductase
F:236-372/Domain: cytochrome P450 homology <CYP>

Query Match 21.2%; Score 450; DB 1; Length 402;
Best Local Similarity 32.4%; Pred. No. 3.5e-24;
Matches 128; Conservative 64; Mismatches 179; Indels 24; Gaps 9;

Qy 13 PAFDFKPPAGYADPPPAIRLEATPIFYWDGSRVNLTRYHDVAVFDP-----ERF 67
Db 8 PELVDPDYDFHEDPPYVYRLRDEADLVNNEERNFVAVRSHDVLQGFSDTALSNAV 67
Qy 68 AVSREWESSAEVSAIPELSDMKKYGFGLPREDHARVAKLVNPFSTRAIDLRAEIQ 127
Db 68 GVS---LDPSSRTSEAVKWS-----MLANDDPALHMKRTLVSGFPRKRIELEPVYL 118
Qy 128 RTVDQDLARSGOEFVDVVAEGIPMRASIALKVPACDEKFRFGATRALAGVGLV 186
Db 119 ELARHSDSAQTESFDDVAFAGLPMVDVSELIGVDTDRARRALADAVHREDEVA 178
Qy 187 LVPODEETKLVASVTEGLALLHDVLDERRRNPLENDVLTMLQAEADGSRSLSTKELVA 246
Db 179 DVP-----PRMAASI-ELMRYVADLIEFRRRP-ANNTLSALLAEIDGRLSDQEIWA 231
Qy 247 LVGAIAGTGTITYLIAFAVLNLRSPBEALVELYKABGRLMNLDEVLPRDNLIRIGTV 306
Db 232 FLFLVAVIGNETTTKLLNNAVYMAAHHGQALAFVADHSRIPMWEETLRVDTSSQI-LA 290
Qy 307 RPARQDLEYCASIKKEMVFLIPSAIRDGTFSRDPVDRDGTASLAVGRCFVPCGV 365
Db 231 RTVADLDLVTYTTIPESGVLLLPESANRDRVRDPDPDYIRIGRICKLVSFGSGAIFC 350
Qy 366 PGVSIARLEAEIAGVTIRRRPEMKLKETPVFGYH 400
Db 351 LGAHLARWEARVALGALRRIRNVEVDDNVVRYH 385

RESULT 11

C87170
Probable cytochrome P450 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: C87170
R:Coile, S.T.; Biglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthey, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Butler, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: C87170
A:Status: preliminary

THIS PAGE BLANK (USPTO)

52

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:34:03 ; Search time 23.9103 seconds
(without alignments)

2772.494 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124
Sequence: 1 MTOEQANQSETPAFDFKPF.....HPAFNIESLVILKPSKAG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters: 587654

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2124	100.0	419	US-10-014-717-8	Sequence 8, Appli
2	505.5	23.8	404	US-10-214-446-50	Sequence 32, Appli
3	498.5	23.5	392	US-10-214-446-32	Sequence 50, Appli
4	497.5	23.4	388	US-10-156-761-13776	Sequence 13776, A
5	495.5	23.3	408	US-10-214-446-4	Sequence 4, Appli
6	491	23.1	475	US-10-145-415-16	Sequence 16, Appli
7	488.5	23.0	475	US-10-145-415-22	Sequence 22, Appli
8	488	23.0	416	US-09-861-289-39	Sequence 39, Appli
9	488	23.0	416	US-09-860-846-39	Sequence 39, Appli
10	488	23.0	416	US-09-988-384B-39	Sequence 39, Appli
11	488	23.0	416	US-09-836-821-39	Sequence 39, Appli
12	485.5	22.9	470	US-10-145-415-6	Sequence 6, Appli
13	485.5	22.9	430	US-10-145-415-30	Sequence 30, Appli
14	484	22.8	416	US-09-793-708-18	Sequence 18, Appli
15	484	22.8	416	US-10-201-365-13	Sequence 13, Appli

16	484	22.8	416	US-10-160-539-18	Sequence 18, Appli
17	483.5	22.8	425	US-10-214-446-20	Sequence 20, Appli
18	483	22.7	399	US-10-156-761-9914	Sequence 9914, Ap
19	480.5	22.6	430	US-10-145-415-95	Sequence 95, Appli
20	475.5	22.4	430	US-10-145-415-8	Sequence 8, Appli
21	475.5	22.4	430	US-10-145-415-12	Sequence 12, Appli
22	475.5	22.4	430	US-10-145-415-20	Sequence 20, Appli
23	466	21.9	432	US-10-145-415-10	Sequence 10, Appli
24	463.5	21.8	430	US-10-145-415-32	Sequence 32, Appli
25	452.5	21.3	416	US-10-156-761-14997	Sequence 14997, A
26	452.5	21.3	430	US-10-145-415-18	Sequence 18, Appli
27	-451.5	21.3	429	US-10-145-415-14	Sequence 14, Appli
28	450	21.2	393	US-10-156-761-9525	Sequence 9525, Ap
29	450	21.2	405	US-10-156-761-12073	Sequence 12073, A
30	448	21.1	430	US-10-145-415-4	Sequence 4, Appli
31	448	21.1	430	US-10-145-415-28	Sequence 28, Appli
32	447	21.0	430	US-10-145-415-24	Sequence 24, Appli
33	447	21.0	457	US-10-156-761-11073	Sequence 11073, A
34	443.5	20.9	430	US-10-145-415-36	Sequence 36, Appli
35	443	20.9	399	US-10-314-657-37	Sequence 37, Appli
36	435.5	20.5	426	US-10-145-415-34	Sequence 34, Appli
37	435.5	20.5	430	US-10-145-415-2	Sequence 2, Appli
38	435	20.5	415	US-10-214-446-56	Sequence 56, Appli
39	433.5	20.4	399	US-10-156-761-7959	Sequence 7959, Ap
40	431	20.3	397	US-10-156-761-14715	Sequence 14715, A
41	428	20.2	415	US-10-214-446-46	Sequence 46, Appli
42	427.5	20.1	416	US-10-156-761-7958	Sequence 7958, Ap
43	425.5	20.0	402	US-10-205-032-8	Sequence 8, Appli
44	425	20.0	410	US-10-205-032-6	Sequence 6, Appli
45	418	19.7	404	US-10-214-446-40	Sequence 40, Appli

ALIGNMENTS

RESULT 1				
US-10-014-717-8				
; Sequence 8, Application US/10014717				
; Publication No. US20020192778A1				
GENERAL INFORMATION:				
; APPLICANT: Schupp, Thomas				
; APPLICANT: Ligon, James				
; APPLICANT: Molnar, Istvan				
; APPLICANT: Zirkle, Ross				
; APPLICANT: Cyr, Devon				
; APPLICANT: Goerlach, Joern				
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHLONES				
; FILE REFERENCE: 4-30582A				
; CURRENT APPLICATION NUMBER: US/10/014,717				
; CURRENT FILING DATE: 2001-11-13				
; PRIOR APPLICATION NUMBER: US/09/335,409				
; PRIOR FILING DATE: 1999-06-17				
; NUMBER OF SEQ ID NOS: 30				
; SOFTWARE: PatentIn Ver. 2.0				
; SEQ ID NO 8				
; LENGTH: 419				
; TYPE: prt				
; ORGANISM: Sorangium cellulosum				
US-10-014-717-8				
Query Match				
Best Local Similarity 100.0%; Pred. No. 2.2e-198;				
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	MTOEQANQSETPAFDFKPFAGYAEDEPPAIEERRENTPIFYWDEGSGWVTRHYDVA	60	
DB	1	MTOEQANQSETPAFDFKPFAGYAEDEPPAIEERRENTPIFYWDEGSGWVTRHYDVA	60	
QY	61	VRDRFPAVSREWSAAYSAIPELSDMKYKFGLPPEEDHARVKLVNPSFTSRAD	120	
DB	61	VRDRFPAVSREWSAAYSAIPELSDMKYKFGLPPEEDHARVKLVNPSFTSRAD	120	
QY	121	LRAEIOFTVQDLARSGOEEFDVVDYAEIGPMRAISALUKVAECDEKERRRGSATA	180	

Db	121	LTRAEIGRTYDQLLDARSGGEFFDVRYDAEGIPRAAISALILKVAPECDKFRFRGSATA	180
Qy	181	RAIVGCVIPVQDDEETITLVASVTEGALATHVDLDRRRPLQENDVLTMLTAAEADSGRLS	240
Db	181	RAIVGCVLPVQDDEETITLVASVTEGALATHVDLDRRRPLQENDVLTMLTAAEADSGRLS	240
Qy	241	TKELVALVGAIIAAGDTTIIYLIIAFAVINILRSPEALBIVKAPGIMRNALDEVLRPDI	300
Db	241	TKELVALVGAIIAAGDTTIIYLIIAFAVINILRSPEALBIVKAPGIMRNALDEVLRPDI	300
Qy	301	LRIGTVAFARQDLEYCGASIKKGEWFFLIIPSLARDGVFSRQDVFVDRDQTGASLAYGR	360
Db	301	LRIGTVAFARQDLEYCGASIKKGEWFFLIIPSLARDGVFSRQDVFVDRDQTGASLAYGR	360
Qy	361	GPVHCPCVSLARLEAETIAVGTIPRRPEPKIKETVFEVGNAPFRNIEELNIIILKPSKXG	419
Db	361	GPVHCPCVSLARLEAETIAVGTIPRRPEPKIKETVFEVGNAPFRNIEELNIIILKPSKXG	419

```

RESULT 2
US-10-214-446-50
; Sequence 50, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 404
; TYPE: prt
; ORGANISM: Bacterial
US-10-214-446-50

```

[illegible]

Oy 369 SLARLEAIAVGTFRRFPPEMKL-ETPVFGIHDAFRNIESNLVILKPS 416
:|||:::|:::|:::|
Db 354 ALAELEGVSRFTLDRLPGLRTHPERIPAPMLRGFASDLDGFRS 402

```

RESULT 3
US-10-214-446-32
; Sequence 32, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-32

```

```

Query Match 23.5% Score 498.5; DB 12; Length 392;
Best Local Similarity 33.8%; Pred. No. 6,66-40;
Matches 127; Conservative 64; Mismatches 164; Indels 21; Gaps 5;

QY 20 FAPGYAEDPPFAIRLRLEATPIF--YWDEGRSWLTRYHDVSAVFRDERFAVSREEMES 76
DB 9 YGADFTANFYVYAKLREAGPVHVEVRMPDGFQFWLTVVGHEGRAALADPRLA-----K 61
QY 77 SAEVSSAIPELSDMKKTGGLFGLPREDIARVAKLVNPFSTSAIDLLRAEIORVDQLDA 136
DB 62 SPVSIVGVPPEEDIDGVHLLAADAPDHTRLRLVTGEFTGRRVGLRRRIQDLTTLELADA 121
QY 137 RSGOEFPVVDVYAEGLPMRAISALLKVPAECDECFRRFGSATRARALGVGLVPQVDEETK 196
DB 122 MEPRGRADLVAPAFPLPIYICELLCVPADRDRTFRMSQGL-----VTPGDQE-- 172
QY 197 TLVASVTEGALLHDVLDERRRNLENDVLTMLLOAE-DCSRUSTEYELVAGAIINAG 255
DB 173 -FGQAMVFAVLDALIREAKRAAGPTDILSALTITARAEDDRLSGPELRAMAYLLIING 231
QY 256 TDTTITVLIAPVNLKRSPALBELVKAEPGLRMAALDEVAFDNILIRGTFRFAQDLEY 315
DB 232 HETTVNLILANVRNILLTHPEQLAALRRAPDILLDGTIESLKYDDGVEGTGTRFREAVTI 291
QY 316 CGASIKKGEVFLIIPSAIRDGTVFSRPDVEDVARDTASLAYSARGPVCVGSVILARLEA 375
DB 292 GGREITAGQYLVVGIAGLDRDPAPFPPDPDRDIRDRTGHLAFGHGHIYCGAPLARLEG 351
QY 376 EIAVGTIFRRPEMKL 391
DB 352 RIARLTLLDRPPDELL 367

RESULT 4
US-10-156-761-13776
; Sequence 13776, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN

```

```

; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13776
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13776

```

```

Query Match      23.4%; Score 497.5; DB 15; Length 388;
Best Local Similarity 34.0%; Pred. No. 8.1e-40;
Matches 136; Conservative 59; Mismatches 144; Indels 61; Gaps 10;

Qy 21 APGAEDPFAIERREKATPIFY---WDEGKSWLTRYHDVSAVFRDERFAVSREEMESS 77
Db 4 AEGGLADYAYVDRLDTPAPVHRIAGTGKPMALVTRYDDV-----RE----- 46
Qy 78 AEYSAPISLDMKXYGFLG-----LPP-----EDHARVRKLVNPSFTSRAID 120
Db 47 ---GLANPLSLDKKHALPGNYRGIALPPLDANILNMDAPDHTIRKLVGKRAFTLRVE 103
Qy 121 LLRAEIQRTVDLIDARSGOEEDVDVRYAEGIPMAISALIKVPAECDEKFRFGSATA 180
Db 104 QLRREVRTARHLIDALGTHGSTDLIASYAAPLPITVICDLGVE---DEHRRD----- 155
Qy 181 PALGVGV-----POVDEETKTLVASTEGIALIHVDERRRNPLENDVITMLQAEA 234
Db 156 RAMTDPVTPPPARDVARE-----SVVSLGFFFTGLADKRNKPAD-DLISDLIAVOE 208
Qy 235 DGSRLSTKEIVLVAGAIITAGDTTIVLIAFAVLNLLRSPLELVKAEPLGRVALDEV 294
Db 209 EGDRIETBELMSLAPLIPAGENTVHLIGNVALLLHPPEGLALREDPARLPVAVGEF 268
Qy 295 LRFNUIRIGTVRFARQDLEYGASIKKGEWFLIIPSLARDGTFSRSDVFDVARDTGA 354
Db 269 ARYEGPALAIRPFVRYVTIGGVTPAGEYTLISLSANRDPSPFPDRDLGRDAAG 328
Qy 355 SLAYGRGPHVCGVSLALEIAIVGTFRRPEPKLKET 394
Db 329 HIALGHGVHCGAPLARLETEVALALALREFPDIALAET 368

```

```

RESULT 5
US-10-214-446-4
; Sequence 4, Application US/10214446
; Publication No. US20030180742a1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4

```

```

; LENGTH: 408
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-4

```

```

Query Match      23.3%; Score 495.5; DB 12; Length 408;
Best Local Similarity 32.3%; Pred. No. 1.4e-39;
Matches 122; Conservative 72; Mismatches 157; Indels 27; Gaps 8;

Qy 27 DDPFAIERRE-----ATPIFYWDEGKSWLTRYHDVSAVFRDERFAVSREEMESSA 78
Db 21 NPYLFEIRREHGVQVRNLPVL-----EVMWGTGDEAVAAALTDPLLS-SSPVGVNGL 74
Qy 79 EYSAPISLDMKXYGFLGPEEDHARVRKLVNPSFTSRAIDLLRAEIQRTVDLIDARS 138
Db 75 EEBMAHQERTIVLWASMLVANGEDHTRLRNVLSAFTARVEQLAPRQAHTDALDVA 134
Qy 139 QOEEDVDVRYAEGIPMAISALIKVPAECDEKFRFGSATAHALGVGV---POVDEETK 196
Db 135 ARGSDLVSEFALPLPMANVLSDLIGIPAEQPDPPAR-----LAVGLIMPNTPERLA 186
Qy 197 TLVAVTEGALLHNDVLDERRRNPLENDVITMLQAEADGSRISTKEIVLVAGAIAGT 256
Db 187 KGAPRADELTFEEPLIAQRKEP-KODLLSALCAQAE-ERISDRLEITAMTILTLTLAGH 244
Qy 257 DTTIYLIAFAVLNLLRSPLELVKAEPLGRNALDEVLPDNIIRIGTVRFARQDLEYC 316
Db 245 ETTASLIANGVHALLRHEQGFATLRDPSLLPGAIIBELRTEGPGVSRGVAFPTDPYEIG 304
Qy 317 GASIKKGEWFLIIPSLARDGTFSRSDVFDV-RDRTGASLAYGRGPHVCGVSLARLEA 375
Db 305 GVTVPAGEMIIIGLAANRDPARYDRPDILDVAREVPOQLAFGHVFCIGAPLARAEA 364
Qy 376 EIAVGTIFRRPEPKLKE 393
Db 365 RIAIGTLRRFPDLRLAD 382

```

```

RESULT 6
US-10-145-415-16
; Sequence 16, Application US/10145415
; Publication No. US20030068788a1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; TITLE OF INVENTION: Methods and Compositions for Making Emamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; PRIOR FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Streptomyces albofaciens
US-10-145-415-16

```

```

Query Match      23.1%; Score 491; DB 15; Length 475;
Best Local Similarity 29.6%; Pred. No. 4.8e-39;
Matches 129; Conservative 83; Mismatches 156; Indels 68; Gaps 11;

Qy 6 ANOSETKPAIDPKFAPGVA-----EDPFAIERREKATPIFY---YMEEG 47
Db 34 ASPDTSBATDRTTLTSYVGLHGPENLMEPELLDNFYGTGLREDAPLVRAAFIDDS 93
Qy 48 RSWLTRYHDVSAVFRDERFAVSREEMESSAEYSAPISL-----DMKXYGFLGFP--- 99

```

```

Db 94 PIMVTRFVNEVWMDQF-----VNNPTLVPGGADQDPRALIELFGIPEDL 143
QY 100 -----PEDHARVKLVNPSFTSRALDLPAEIQRTVDQLLDARSQOE--FD 144
Db 144 AYLDTITLTPDPDHTLRRLVSRAFTARRIODLPRVERITDELIALUPDAEAGVAD 203
QY 145 VVRDVAEGIPMARISALLVPAECDEKFRFGSATRALGVGVPOVDERTKLVASYTE 204
Db 204 LVEHFVYPIPIYVICELVGDDEDRALMRFPGADLA-----SLNP-----KRIGATMB 252
QY 205 GUALHDVDERRRNPLENDVLTMLQA-EADGSRLSKTELVALVGAIIAAGTDTTYLI 263
Db 253 MTAHIEVIDERAD-LRDLLSGLIRAQODDGRISDVEMVTLVTLVAGHETTAHI 311
QY 264 ARAVNLASPEALEVKAPEGMRNALDEVLFNDILIRIGVRFARODLEYGASIKKG 323
Db 312 SNGTLLALTHPDORRLIDEDPALLPRAVHELMKCGPIQATQRLYAMEDETVAGVQROG 371
QY 324 EMVFLIPSLARDCGVFSRPDVFVRDGA-----SLAYGRGPHVCGVSLARLEAIV 379
Db 372 EALMSLVANNDPRHYTGPERLDLTPQAGRAEDHVGFGHMYCLGASLARQEAHVAY 431
QY 380 GTIFRRPEPKLKETP 395
Db 432 GKULTRYPDALALTLP 447

```

RESULT 7

```

US-10-145-415-22
; Sequence 22, Application US/10145415
; Publication No. US20030068788A1
GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zitzke, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Enamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 475
; TYPE: PRF
; ORGANISM: Streptomyces
US-10-145-415-22

```

```

Query Match 23.0%; Score 488.5; DB 15; Length 475;
Best Local Similarity 29.7%; Pred. No. 8.5e-39;
Matches 131; Conservative 83; Mismatches 158; Indels 69; Gaps 12;

```

```

QY 2 TOSQANOSETK-PAPDFKPPAPGYA-----EDPPAIERLEATPIF--- 42
Db 29 TPEAAATSSDPTTYPATDRTTLPSYVGLHPGEPLMEPELLDNDYTGVTGREAQPIVRRAR 88
QY 43 YMEGGSWVLTTRYHDVSAVFRDERFAVSREMESSAAYSAPISLSMKK-----YGLRG 97
Db 89 FIDDSPIWLTFRDVVREVWDRQF-----VNNPTLVPGIGADKOPRARLIELFG 138
QY 98 LP-----PEDHARVKLVNPSFTSRALDLPAEIQRTVDQLLDARSQOE 142
Db 139 IPBDLTPYLADTILTSPPHTTLRLKRVLSAFTARRIQLRPRVEQITDALLERLPDAE 198
QY 143 ---FDVVRDVAEGIPMARISALLVPAECDEKFRFGSATRALGVGVPOVDERTKTLV 199
Db 199 DGAVDVEHFVYPIPIYVICELVGDDEDRALMRFPGADLA-----SLNP-----KRIG 247

```

```

QY 200 ASVTEGALLHDVDERRRNPLENDVLTMLQA-EADGSRLSKTELVALVGAIIAAGTDT 258
Db 248 ATPEMIAHIEVIDERRA-ALRDLLSGLIRAQODDGRISDVEMVTLVTLVAGHET 306
QY 259 TYLIAFAVNLARSPEALEVKAPEGMRNALDEVLFNDILIRIGVRFARODLEYGCA 318
Db 307 TAILHISNGTLLALTHPDORRLIDEDPALLPRAVHELMKCGPIQATQRLYAMEDETVAGV 366
QY 319 SIKKGMVFLIPSLARDCGVFSRPDVFVRDGA-----SLAYGRGPHVCGVSLARLE 374
Db 367 QVROGBALMFSLVANNDPRHYTGPERLDLTPQAGRAEDHVGFGHMYCLGASLARQ 426
QY 375 AEIAGTIFRRPEPKLKETP 395
Db 427 AEVAGKULTRYPDALALTLP 447

```

RESULT 8

```

US-09-861-289-39
; Sequence 39, Application US/09861289
; Patent No. US2002110897A1
GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FaastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRF
; ORGANISM: Streptomyces venezuelae
US-09-861-289-39

```

```

Query Match 23.0%; Score 488; DB 10; Length 416;
Best Local Similarity 30.5%; Pred. No. 7.7e-39;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

```

```

QY 1 MTOEQANOSETK-PAPDFKPPAPGYAEDPPAIERLEATPI--FTWDEG-RSMVLTTRYHD 57
Db 1 MRRTOGGTTASPPVLDLGLAGDFAADPYPTVARRLAESGPAAHVRTPBGDEWLVVGIDR 60
QY 58 VSAVFRDERFAVSREMESSAAYSAPISLSMKKGYLGLPPEDHARVKLVNPSFTSR 117
Db 61 ARAVNLADPRFS---KDMRNS---TTPDLTEAALNHNMLESDPPRHTRLRKLVAREFTWR 114
QY 118 AIDLRAEIQRTVDQLLDARSQOE--FDVVRDVAEGIPMARISALLVKAECDEKFRF 175
Db 115 REVELRPRVOEIVDGLVDMLAAPDGRADLMESLAMPPLITVISELGVPEEDRAAFRW 174
QY 176 GSNTARALGVGVPOVDERTKLVASVTEGLLHDVDERRRNPLENDVLTMLQAED 235
Db 175 TDAF-----VPPDPAQAQTMAEMSGYLSRL--IDSKGQDGEDLSALVRTSDSD 224
QY 236 GSRLSKTELVALVGAIIAAGTDTTYLIAFAVNLARSPEALEVVAEBGLMRNALDEV 295
Db 225 GSRUTSEELLGAHAILLVAGHETTVNLINGWYALLSHRDQALALRADMTLDGAVEMVL 284
QY 296 REDNILIRIGTVRFARODLEYGASIKGMVFLIPSLARDCGVFSRPDVFVRDRTGAS 355
Db 285 RRGVPESATYRPPVFDLDTGTVIPAGDTVLVADARTEPRFPDPFRFDRDRTAGH 344
QY 356 LAYGRPHVCGVSLARLEAIVAGTIFRRPEPKLKETP---VFGYHAPFNIESTLV 411
Db 345 LAFGHIHFCIGAPLARLEARIAVRLALRCPDLADVSGELIWMYPNMIRKALKAPI 403

```



```

RESULT 9
US-09-860-846-39
; Sequence 39, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FaestSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-39

```

```

Query Match      23.0%; Score 488; DB 10; Length 416;
Best Local Similarity 30.5%; Pred. No. 7.7e-39;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

```

```

QY 1 MTQEQANSETPADPFKFPAGYADDPFAIERLEATPI--FYWDEG-RSMVLTTRYD 57
DB 1 MKRTQGGTTASPPVLDLGLGODFADPYPTAARLAEGBPAHVRTPEDDEVWLVGXYDR 60
QY 58 VSAVPRDERFAVSREEMESSAESSAIPELSDMKKYGFLGPPEDHARVRKLVNPSFTSR 117
DB 61 ARAVLADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPHTRLRKLVAREFTWR 114
QY 118 AIDLRAEIQRTVDQLLDARSGOE--FDVVRDYAEGIPMRAISALLKVPACDEKFRFR 175
DB 115 RVELLRPRVOEIVDGLVDMMLAAPGRADLMESLAMPLEITVISELGVPEPDRAAFRW 174
QY 176 GSATRALGVGLVPOVDEETKTLVASVTEGALLHDVLDERRRNPENDVLTMLQAEND 235
DB 175 TDAF-----VFPPDPAQAOTAMAMESGYLSRL---IDSKRGQGGEDLLSLVRTSDSD 224
QY 236 GSRLSTKEIVLVGAIITAGTDTTYLIAFAVLNLRSPALRLVKAEGCLMRNALDEV 295
DB 225 GSRLTSEELGMAHILVAGHETTVNLANGMYALLSHPDQLAALRADTLLDGAVEEML 284
QY 296 REDNLRIGTVRFARQDLEYCGASIKKGMVFLIPSLARDGVFSRPPVFPVRRDTGAS 355
DB 285 RREGPVESATYRFPVPEVDLDTVIPAGDTVLVVLADAHRTBERPDPHFRFIRDTTACH 344
QY 356 LAYGCPHYCPGVSLARLEAEIAGVTIFRRPEMKIKETP---VGYHPAFRNIESLVN 411
DB 345 LAFGIGHFCIGAPLARLEARIAVNALLERCPDLADVSPGELVWYPNMIGLKALPLI 403

```

```

RESULT 10
US-09-988-384B-39
; Sequence 39, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/143398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537

```

```

; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-39

```

```

Query Match      23.0%; Score 488; DB 11; Length 416;
Best Local Similarity 30.5%; Pred. No. 7.7e-39;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

```

```

QY 1 MTQEQANSETPADPFKFPAGYADDPFAIERLEATPI--FYWDEG-RSMVLTTRYD 57
DB 1 MKRTQGGTTASPPVLDLGLGODFADPYPTAARLAEGBPAHVRTPEDDEVWLVGXYDR 60
QY 58 VSAVPRDERFAVSREEMESSAESSAIPELSDMKKYGFLGPPEDHARVRKLVNPSFTSR 117
DB 61 ARAVLADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPHTRLRKLVAREFTWR 114
QY 118 AIDLRAEIQRTVDQLLDARSGOE--FDVVRDYAEGIPMRAISALLKVPACDEKFRFR 175
DB 115 RVELLRPRVOEIVDGLVDMMLAAPGRADLMESLAMPLEITVISELGVPEPDRAAFRW 174
QY 176 GSATRALGVGLVPOVDEETKTLVASVTEGALLHDVLDERRRNPENDVLTMLQAEND 235
DB 175 TDAF-----VFPPDPAQAOTAMAMESGYLSRL---IDSKRGQGGEDLLSLVRTSDSD 224
QY 236 GSRLSTKEIVLVGAIITAGTDTTYLIAFAVLNLRSPALRLVKAEGCLMRNALDEV 295
DB 225 GSRLTSEELGMAHILVAGHETTVNLANGMYALLSHPDQLAALRADTLLDGAVEEML 284
QY 296 REDNLRIGTVRFARQDLEYCGASIKKGMVFLIPSLARDGVFSRPPVFPVRRDTGAS 355
DB 285 RREGPVESATYRFPVPEVDLDTVIPAGDTVLVVLADAHRTBERPDPHFRFIRDTTACH 344
QY 356 LAYGCPHYCPGVSLARLEAEIAGVTIFRRPEMKIKETP---VGYHPAFRNIESLVN 411
DB 345 LAFGIGHFCIGAPLARLEARIAVNALLERCPDLADVSPGELVWYPNMIGLKALPLI 403

```

```

RESULT 11
US-09-836-821-39
; Sequence 39, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FaestSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-836-821-39

```

```

Query Match      23.0%; Score 488; DB 11; Length 416;
Best Local Similarity 30.5%; Pred. No. 7.7e-39;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

```

```

QY 1 MTQEQANSETPADPFKFPAGYADDPFAIERLEATPI--FYWDEG-RSMVLTTRYD 57
DB 1 MKRTQGGTTASPPVLDLGLGODFADPYPTAARLAEGBPAHVRTPEDDEVWLVGXYDR 60
QY 58 VSAVPRDERFAVSREEMESSAESSAIPELSDMKKYGFLGPPEDHARVRKLVNPSFTSR 117

```

```

Db      61 ARAVALDPRFS---KOWRNS---TTPLEAEALNNMLESODPRTRLAKLVAREFTNR 114
QY      118 AIDLRAEIQRTVDOLLDRASGOEE--FDVVRDYAGCIPMRASISALLKVAECDEKFRFP 175
Db      115 RVELLRPRVQELVDGLVDMLAAPDRADIMESLAMPITVISELGGVEPPRAAFRW 174
QY      176 GSATARALGVLPOVDEEFTKTVASVTEGLALLHVDLDERRRNPLENDVLTMLQAEAD 235
Db      175 TDAF-----VPPDDPAQQTMAEMSGYLSRL---IDSKROGDEDLISALVTSDED 224
QY      236 GSRISTKELVALGAIIAAGTDTTTLIAFAVNLILRSPEALEVAEPGLKRNALDEVL 295
Db      225 GSRITSEELGVMHILLVAGHETTVMILANGMYALISHPQDLAALRADMTLLDGAVEML 284
QY      296 RFDNIRIGTVRARADLEVCASIKKGEVFLILISALRDGVFSRPFVDRDPTGAS 355
Db      285 RYEGPESATYRPPVPEVDLGTIVIPAGDVLVVLADAHETPERFPDPRHFDIRDPYAGH 344
QY      356 LAYGRGPHVCPGVSLARLSAEIAGVTIFRREPMMKLETP---VFGYHAFRNIESLNV 411
Db      345 LAFGHGHCIGAPLARLEKARIVARALLERCPDLADVSGELVWTPNPIRGKALPI 403

```

RESULT 12

```

US-10-145-415-6
; Sequence 6, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zirkle, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Enamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; PRIOR FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 470
; TYPE: PRT
; ORGANISM: Streptomyces rimosus
US-10-145-415-6

```

```

Query Match      22.9%; Score 486.5; DB 15; Length 470;
Best Local Similarity 30.6%; Pred. No. 1.3e-38;
Matches 124; Conservative 81; Mismatches 147; Indels 53; Gaps 10;

QY      22 PGYAEPPPAIERLEKATPIF---YWDGRSWLTRYHVSAAVFRDERFAVSEBMESSA 78
Db      60 PELLENPYTCYGLRREGAPLVRAAFIDSPILWLTREVVRVDRF-----V 109
QY      79 EYSSAIPELSDMKK---YGLFGLP-----PEDHARVRLVNPSTFSTRA 118
Db      110 NNPTLVPGIGADDDPRARLLELFGIPEDLAPYLTDNLITSDPPDHTLRRLVSRAPYARR 169
QY      119 IDLLRAEIQRTVDOLLDRASGOEE--FDVVRDYAGCIPMRASISALLKVAECDEKFRFP 175
Db      170 IQDLRPRVERITTELERLPDHAEDGVVDLVEHFAVPLPTVICELVGIDEBDRAMRRF 229
QY      176 GSATARALGVLPOVDEEFTKTVASVTEGLALLHVDLDERRRNPLENDVLTMLQAEAD 235
Db      230 GADLA-----SLNP-----KRIGATPEMISHHEIDERRA-ALADDILSGIIRQDD 277
QY      235 DGSRLSTKELVALGAIIAAGTDTTTLIAFAVNLILRSPEALEVAEPGLKRNALDEV 294
Db      278 DGRISDVENVTIVLTLVLAGHETTALHISNGTIALTLTHPDQRLIDEDBALLPRAVHEL 337

```

```

QY      295 LRFDNIRIGTVRARADLEVCASIKKGEVFLILISALRDGVFSRPFVDRDPTGA 354
Db      338 MWCCEGIQTOURLRYALDEVEVAGVQVROGEALMFLSVANNDPRHVTGGERLDLTRQPG 397
QY      355 ---SLAYGRGPHVCPGVSLARLSAEIAGVTIFRREPMMKLETP 395
Db      398 RAEDHVGFGHGMHYCIGASLARQEAFAVAYGKLLTRFPDLALALTP 442

```

RESULT 13

```

US-10-145-415-30
; Sequence 30, Application US/10145415
; Publication No. US20030068788A1
; GENERAL INFORMATION:
; APPLICANT: Buckel, Thomas Gunter
; APPLICANT: Hammer, Philip Eugene
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ligon, James Madison
; APPLICANT: Molnar, Istvan
; APPLICANT: Pachlatko, Johannes Paul
; APPLICANT: Zirkle, Ross Eric
; TITLE OF INVENTION: Methods and Compositions for Making Enamectin
; FILE REFERENCE: SYN-117 109846.312
; CURRENT APPLICATION NUMBER: US/10/145,415
; CURRENT FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/291,149
; PRIOR FILING DATE: 2001-05-16
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Streptomyces lydicus
US-10-145-415-30

```

```

Query Match      22.9%; Score 485.5; DB 15; Length 430;
Best Local Similarity 31.4%; Pred. No. 1.4e-38;
Matches 129; Conservative 81; Mismatches 150; Indels 51; Gaps 13;

QY      22 PGYAEPPPAIERLEKATPIF---YWDGRSWLTRYHVSAAVFRDERFAVSEBMESSA 78
Db      25 PALIGDPFAGYGALRQGGVVRGRFVDSPPVFTREEVREVLKDRF---RNPVSSA 81
QY      79 EYSSAIPELSDMKK---YGLFGLPPE-----DHARVRLVNPSTFSTRAIDLL 122
Db      82 --PDADPEDTPLSRLDMMGFPBEHLRVYLIGSLINNDADPHTRRLVSRAPYARRITDL 139
QY      123 RAEIQRTVDOLLDRASGOEE--FDVVRDYAGCIPMRASISALLKVAECDEKFRFGSAT 179
Db      140 RPRVAGIADDELLARLEHAEEDGVVDLIQHFAVPLPTVICELVGIDEBDRPQWRWGADL 199
QY      180 ARAALGVLPOVDEEFTKTVASVTEGLALLHVDLDERRRNPLENDVLTMLQAEADGSR 238
Db      200 ----SLNP-----DRMSRSSPAMIDH---IHETIAARR-ALTDLLSELIRTHDDGSR 247
QY      239 LSTKELVALGAIIAAGTDTTTLIAFAVNLILRSPEALEVAEPGLKRNALDEVLRFP 298
Db      248 LSDVENVTIVLTVLAGHETTALHISNGTIALTLTHPDQRLKDDPALLPRAVHELMRNC 307
QY      299 NLRIGTVRARADLEVCASIKKGEVFLILISALRDGVFSRPFVDRDPTGA---- 354
Db      308 GPVHMTQLRYAABDVDELAVGRIRKGDVQVLIVSANRDPRHYTEPRDLTRHPAGHAEN 367
QY      355 SLAYGRGPHVCPGVSLARLSAEIAGVTIFRREPMMK-----LKEPVPFG 398
Db      368 HVGFGHAGHCIGATLAKOEGEVALGALLRHPEELSLAVALALERTPVPG 418

```

RESULT 14

```

US-09-793-708-18
; Sequence 18, Application US/09793708
; Publication No. US20030104597A1

```

```

; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PR
; ORGANISM: Streptomyces venezuelae
US-09-793-708-18

```

```

Query Match      22.8%; Score 484; DB 11; Length 416;
Best Local Similarity 30.6%; Pred. No. 1.9e-38;
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQETKAPFKPAGYADPPAIREATPI--FYWDEG-RSMVLTIRYHDSAV 61
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 QGGTASPPVLDLGLGQDFADPPYPTARLAEGBAHRVRTPEGDVWLVGVDRAAV 64
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 FDERFAVSREEMESSAEYSSAIPELSDMKYGLFGLPPEDHARVRLVNPSTSAIDL 121
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 LADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPRTRKLKLVAREFTMRVRL 118
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 LRAEIQTVDOLLARSQGE--FDVVRDYAEGIPMAISALLKYPACDEKFRFGSAT 179
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 LRPVQEIYDGLVDAMLAPDGRADLMESLWPLPTIVISELGVPEPRAAFRWTDAP 178
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 AALGVGLVPOVDEETKTLVASVTBGLALHVDLERRRNPLENDVLTMLQAEADGSL 239
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 -----VFPDDPAQAQTAMAEWSGYLSRL--ISKRGQDGEDLLSALVRTSDEGSRL 228
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 STKEIVLVGALIIAGTDTTIVLIAFAVNLRSPEALEVKAEPGLMKNALDEVLRPN 299
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 229 TSEELGMAHILLVAGHETTIVLIANGMVALSHPDQLAALRADMTLLDGAVEMLRYEG 288
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 ILRIGTVFARODLEYCGASIKKGEVFLIPSLALDGTVFSRPDVFDVRDGTGASLAYG 359
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 289 PVESATYRFPVPEVDLGTIVIPAGDTVLVVLADAHRTPERFPDHRFDIRRDAGHLAFG 348
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 RGPVHCPCVSLARLEAIVGTFIRFRPEMKLKEP---VFGYHAPFRNIESLNV 411
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 HGIIHFCIGAPLARLEARIARIALRCPDLALDVSPGELWVYPNMIRGLKALPI 403
    | : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

```

US-10-201-365-13
; Sequence 13, Application US/10201365
; Publication No. US20030148469A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li

```

```

; TITLE OF INVENTION: COMBINATORIAL POLYKETIDE LIBRARIES PRODUCED USING A MODULAR
; FILE REFERENCE: 300622002103
; CURRENT APPLICATION NUMBER: US/10/201,365
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 416
; TYPE: PR
; ORGANISM: Streptomyces venezuelae
US-10-201-365-13

```

```

Query Match      22.8%; Score 484; DB 12; Length 416;
Best Local Similarity 30.6%; Pred. No. 1.9e-38;
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQETKAPFKPAGYADPPAIREATPI--FYWDEG-RSMVLTIRYHDSAV 61
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 5 QGGTASPPVLDLGLGQDFADPPYPTARLAEGBAHRVRTPEGDVWLVGVDRAAV 64
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 62 FDERFAVSREEMESSAEYSSAIPELSDMKYGLFGLPPEDHARVRLVNPSTSAIDL 121
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 65 LADPRFS---KDMRNS---TTPLTEAALNHNMLESDPPRTRKLKLVAREFTMRVRL 118
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 LRAEIQTVDOLLARSQGE--FDVVRDYAEGIPMAISALLKYPACDEKFRFGSAT 179
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 LRPVQEIYDGLVDAMLAPDGRADLMESLWPLPTIVISELGVPEPRAAFRWTDAP 178
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 AALGVGLVPOVDEETKTLVASVTBGLALHVDLERRRNPLENDVLTMLQAEADGSL 239
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 179 -----VFPDDPAQAQTAMAEWSGYLSRL--ISKRGQDGEDLLSALVRTSDEGSRL 228
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 STKEIVLVGALIIAGTDTTIVLIAFAVNLRSPEALEVKAEPGLMKNALDEVLRPN 299
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 229 TSEELGMAHILLVAGHETTIVLIANGMVALSHPDQLAALRADMTLLDGAVEMLRYEG 288
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 ILRIGTVFARODLEYCGASIKKGEVFLIPSLALDGTVFSRPDVFDVRDGTGASLAYG 359
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 289 PVESATYRFPVPEVDLGTIVIPAGDTVLVVLADAHRTPERFPDHRFDIRRDAGHLAFG 348
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 360 RGPVHCPCVSLARLEAIVGTFIRFRPEMKLKEP---VFGYHAPFRNIESLNV 411
    | : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 HGIIHFCIGAPLARLEARIARIALRCPDLALDVSPGELWVYPNMIRGLKALPI 403
    | : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: October 2, 2003, 17:47:43
 Job time : 25.9103 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:33 ; Search time 10.2473 Seconds
(without alignments)
1730.044 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124

Sequence: 1 MTOBOANSEKTPAFDFKPF.....HPAFNIESLNVILKPSXAG 419

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/CTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2124	100.0	419	3	US-09-335-409-8
2	2124	100.0	419	3	US-09-413-814-71
3	2124	100.0	419	4	US-09-568-102-8
4	2124	100.0	419	4	US-09-567-969-8
5	2124	100.0	419	4	US-09-568-480-8
6	2124	100.0	419	4	US-09-568-486-8
7	2124	100.0	419	4	US-09-568-472-8
8	2124	100.0	419	4	US-09-567-899-8
9	488	22.0	416	3	US-09-105-537-39
10	484.5	22.8	437	4	US-09-252-991A-17836
11	484	22.8	416	3	US-09-320-878-18
12	484	22.8	416	4	US-09-141-908-13
13	484	22.8	416	4	US-09-657-440-18
14	459	21.6	403	6	5212296-9
15	421.5	19.8	468	4	US-09-252-991A-32437
16	371	17.5	409	3	US-09-385-028-12
17	371	17.5	409	4	US-09-726-614-12
18	370.5	17.4	399	4	US-08-765-907A-10
19	361.5	17.0	412	1	US-08-102-863-11
20	361.5	17.0	412	5	PCT-US92-10885-11
21	354.5	16.7	406	6	5212296-6
22	325.5	15.3	395	4	US-09-266-965-129
23	305	14.4	422	1	US-08-396-218-2
24	305	14.4	422	1	US-08-760-116-2
25	300	14.1	422	2	US-09-096-982-5
26	300	14.1	422	2	US-08-653-650A-5
27	300	14.1	474	2	US-09-096-982-8

28	300	14.1	474	2	US-08-653-650A-8	Sequence 8, Appli
29	298	14.0	443	2	US-09-096-982-9	Sequence 9, Appli
30	298	14.0	443	2	US-08-653-650A-9	Sequence 9, Appli
31	262.5	12.4	189	4	US-09-679-279-20	Sequence 20, Appli
32	257	12.1	382	4	US-09-320-878-7	Sequence 7, Appli
33	257	12.1	382	4	US-09-141-908-7	Sequence 7, Appli
34	257	12.1	382	4	US-09-657-440-7	Sequence 7, Appli
35	257	12.1	402	3	US-09-105-537-22	Sequence 22, Appli
36	257	12.1	3782	3	US-09-105-537-4	Sequence 2, Appli
37	185.5	8.7	472	2	US-08-622-166A-2	Sequence 2, Appli
38	185.5	8.7	472	2	US-08-622-166A-4	Sequence 4, Appli
39	184	8.7	520	4	US-09-527-073-2	Sequence 2, Appli
40	175.5	8.3	377	4	US-09-679-279-3	Sequence 3, Appli
41	172.5	8.1	513	4	US-09-351-229-4	Sequence 4, Appli
42	170	8.0	444	3	US-09-413-814-95	Sequence 95, Appli
43	170	8.0	444	3	US-09-413-814-96	Sequence 96, Appli
44	164	7.7	504	4	US-09-128-420A-18	Sequence 18, Appli
45	163	7.7	495	1	US-08-532-065B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1									
US-09-335-409-8									
Sequence 8, Application US/09335409									
Patent No. 6121029									
GENERAL INFORMATION:									
APPLICANT: Schupp, Thomas									
APPLICANT: Ligon, James									
APPLICANT: Molnar, Istvan									
APPLICANT: Zirkle, Rose									
APPLICANT: Cyr, Devon									
APPLICANT: Goetzach, Joern									
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES									
FILE REFERENCE: 4-30582A									
CURRENT APPLICATION NUMBER: US/09/335,409									
CURRENT FILING DATE: 1999-06-17									
NUMBER OF SEQ ID NOS: 30									
SOFTWARE: PatentIn Ver. 2.0									
SEQ ID NO 8									
LENGTH: 419									
TYPE: PRT									
ORGANISM: Sorangium cellulosum									
US-09-335-409-8									
Query Match									
Best Local Similarity 100.0%; Pred. No. 1.2e-210;									
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	MTOBOANSEKTPAFDFKPFAGYAEDEPFAIERREATPIFYMEGRSWVTRYHDSA	60						
DB	1	MTOBOANSEKTPAFDFKPFAGYAEDEPFAIERREATPIFYMEGRSWVTRYHDSA	60						
QY	61	VFRDRFVSVSEEMESSAEYSATPELSDMKYGLFGPPEDHARVRLVNSFTSRAID	120						
DB	61	VFRDRFVSVSEEMESSAEYSATPELSDMKYGLFGPPEDHARVRLVNSFTSRAID	120						
QY	121	LIRAEIQTVDQLDARSQGEFVDVRYDAGICIPRAISALIKVPAECDEKFRFGSATA	180						
DB	121	LIRAEIQTVDQLDARSQGEFVDVRYDAGICIPRAISALIKVPAECDEKFRFGSATA	180						
QY	181	RALGVGVPOVDEEKTIVASVTESIALHVLDRRRNPLENDVLTMLQAEAGSRLS	240						
DB	181	RALGVGVPOVDEEKTIVASVTESIALHVLDRRRNPLENDVLTMLQAEAGSRLS	240						
QY	241	TKELVALVGAIIAAGTDTTIIYLIFAVNLRLSPALSLVAEPGLMNADEVLPDNI	300						
DB	241	TKELVALVGAIIAAGTDTTIIYLIFAVNLRLSPALSLVAEPGLMNADEVLPDNI	300						
QY	301	LRIQTVFARADLEVCASIKKGEVFLILPSALRDGTFSRDPVFDVVRDGTGSLAYGR	360						
DB	301	LRIQTVFARADLEVCASIKKGEVFLILPSALRDGTFSRDPVFDVVRDGTGSLAYGR	360						

QY 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419
Db 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419

RESULT 2

US-09-413-814-71
Sequence 71, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Bayer, Stefan
APPLICANT: Bioecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hofle, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
EARLIER FILING DATE: 1999-10-07
NUMBER OF SEQ ID NOS: 107
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 71
LENGTH: 419
TYPE: PR1
ORGANISM: Sorangium cellulosum
US-09-413-814-71

Query Match 100.0%; Score 2124; DB 3; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.2e-210;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAPDFKFPAPGYADPPPAERLEATPIFYWDEGRSMTLTRYHDVSA 60
Db 1 MTOEQANQSETKPAPDFKFPAPGYADPPPAERLEATPIFYWDEGRSMTLTRYHDVSA 60
QY 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
Db 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
QY 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIIPMAISALLKVPACDEKFRFSGATA 180
Db 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIIPMAISALLKVPACDEKFRFSGATA 180
QY 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGSRLS 240
Db 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGSRLS 240
QY 241 TKELVALVGAIIAAGDTTITLLAFVNLILRSPEALELVKABPGIMRNALDEVLPDNI 300
Db 241 TKELVALVGAIIAAGDTTITLLAFVNLILRSPEALELVKABPGIMRNALDEVLPDNI 300
QY 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTVFSRPDVFVDRDGTASLAYGR 360
Db 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTVFSRPDVFVDRDGTASLAYGR 360
QY 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419
Db 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419

RESULT 3
US-09-568-102-8
Sequence 8, Application US/09568102

Patent No. 6346404
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Rose
APPLICANT: Cyr, Devon
APPLICANT: Goelach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,102
PRIOR FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 419
TYPE: PR1
ORGANISM: Sorangium cellulosum
US-09-568-102-8

Query Match 100.0%; Score 2124; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.2e-210;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAPDFKFPAPGYADPPPAERLEATPIFYWDEGRSMTLTRYHDVSA 60
Db 1 MTOEQANQSETKPAPDFKFPAPGYADPPPAERLEATPIFYWDEGRSMTLTRYHDVSA 60
QY 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
Db 61 VFDERFAVSREWESSAESSAIPELSDMKKYGFGLPEDHARVKLVNPSFTSRAID 120
QY 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIIPMAISALLKVPACDEKFRFSGATA 180
Db 121 LRAEIQRTVDQLDARSQGEFDDVVDYAEIIPMAISALLKVPACDEKFRFSGATA 180
QY 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGSRLS 240
Db 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVITMLQAEADGSRLS 240
QY 241 TKELVALVGAIIAAGDTTITLLAFVNLILRSPEALELVKABPGIMRNALDEVLPDNI 300
Db 241 TKELVALVGAIIAAGDTTITLLAFVNLILRSPEALELVKABPGIMRNALDEVLPDNI 300
QY 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTVFSRPDVFVDRDGTASLAYGR 360
Db 301 LRIGTVRFARODLEYCGASIKKGEWFLIPSLRDGTVFSRPDVFVDRDGTASLAYGR 360
QY 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419
Db 361 GPHVCGVSLARLEAEIAVGTIFRRPEMKLKETPVGYPAPRNIESLNVILKPSKAG 419

RESULT 4
US-09-567-969-8
Sequence 8, Application US/09567969
Patent No. 635457
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Rose
APPLICANT: Cyr, Devon
APPLICANT: Goelach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/567,969
PRIOR FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
NUMBER OF SEQ ID NOS: 30

SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 419
TYPE: PR
ORGANISM: Sorangium cellulosum
US-09-567-969-8

Query Match 100.0%; Score 2124; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.2e-210;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEANOSEKPAEDFKFAPGYAEDPPAERLEATPIFYWDEGSMVLTTRYHDSA 60
DB 1 MTOEANOSEKPAEDFKFAPGYAEDPPAERLEATPIFYWDEGSMVLTTRYHDSA 60
QY 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120
DB 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120
QY 121 LIRAEIORTVDQLDARSQOEEDVVRDYAEGIPMRASALKVPAECDEKRRFGSATA 180
DB 121 LIRAEIORTVDQLDARSQOEEDVVRDYAEGIPMRASALKVPAECDEKRRFGSATA 180
QY 181 RALGVLVPOVDEETKTVASVTGELALHDVLDERRNPLENDVLTMLQAEADGSRIS 240
DB 181 RALGVLVPOVDEETKTVASVTGELALHDVLDERRNPLENDVLTMLQAEADGSRIS 240
QY 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALEVKAEPGLMRNALDEVLRFDNI 300
DB 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALEVKAEPGLMRNALDEVLRFDNI 300
QY 301 LRIGTVRFRARODLEFCGASIKKGENVFLLIPALRDGVFSRPDVFVARDTGASLAYGR 360
DB 301 LRIGTVRFRARODLEFCGASIKKGENVFLLIPALRDGVFSRPDVFVARDTGASLAYGR 360
QY 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKETVFGYHPAFRNIESLNVILKPSKAG 419
DB 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKETVFGYHPAFRNIESLNVILKPSKAG 419

RESULT 5
US-09-568-480-8
Sequence 8, Application US/09568480
Patent No. 6355458
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,480
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 419
TYPE: PR
ORGANISM: Sorangium cellulosum
US-09-568-480-8

Query Match 100.0%; Score 2124; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.2e-210;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEANOSEKPAEDFKFAPGYAEDPPAERLEATPIFYWDEGSMVLTTRYHDSA 60
DB 1 MTOEANOSEKPAEDFKFAPGYAEDPPAERLEATPIFYWDEGSMVLTTRYHDSA 60

QY 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120
DB 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120
QY 121 LIRAEIORTVDQLDARSQOEEDVVRDYAEGIPMRASALKVPAECDEKRRFGSATA 180
DB 121 LIRAEIORTVDQLDARSQOEEDVVRDYAEGIPMRASALKVPAECDEKRRFGSATA 180
QY 181 RALGVLVPOVDEETKTVASVTGELALHDVLDERRNPLENDVLTMLQAEADGSRIS 240
DB 181 RALGVLVPOVDEETKTVASVTGELALHDVLDERRNPLENDVLTMLQAEADGSRIS 240
QY 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALEVKAEPGLMRNALDEVLRFDNI 300
DB 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALEVKAEPGLMRNALDEVLRFDNI 300
QY 301 LRIGTVRFRARODLEFCGASIKKGENVFLLIPALRDGVFSRPDVFVARDTGASLAYGR 360
DB 301 LRIGTVRFRARODLEFCGASIKKGENVFLLIPALRDGVFSRPDVFVARDTGASLAYGR 360
QY 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKETVFGYHPAFRNIESLNVILKPSKAG 419
DB 361 GPHVCPGVSILARLEAIEAVGTIFRRPEMKLKETVFGYHPAFRNIESLNVILKPSKAG 419

RESULT 6
US-09-568-486-8
Sequence 8, Application US/09568486
Patent No. 6355459
GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/568,486
CURRENT FILING DATE: 2000-05-10
PRIOR APPLICATION NUMBER: 09/335,409
PRIOR FILING DATE: 1998-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 419
TYPE: PR
ORGANISM: Sorangium cellulosum
US-09-568-486-8

Query Match 100.0%; Score 2124; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.2e-210;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEANOSEKPAEDFKFAPGYAEDPPAERLEATPIFYWDEGSMVLTTRYHDSA 60
DB 1 MTOEANOSEKPAEDFKFAPGYAEDPPAERLEATPIFYWDEGSMVLTTRYHDSA 60
QY 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120
DB 61 VRDRFPAVSREWESSAESSAIPELSDMKYGLFGLPEDHARVKLVNPSFTSRAID 120
QY 121 LIRAEIORTVDQLDARSQOEEDVVRDYAEGIPMRASALKVPAECDEKRRFGSATA 180
DB 121 LIRAEIORTVDQLDARSQOEEDVVRDYAEGIPMRASALKVPAECDEKRRFGSATA 180
QY 181 RALGVLVPOVDEETKTVASVTGELALHDVLDERRNPLENDVLTMLQAEADGSRIS 240
DB 181 RALGVLVPOVDEETKTVASVTGELALHDVLDERRNPLENDVLTMLQAEADGSRIS 240
QY 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALEVKAEPGLMRNALDEVLRFDNI 300
DB 241 TKEVALVGAIIAAGDTTIIYLAFAVNLRSPEALEVKAEPGLMRNALDEVLRFDNI 300

QY 301 LRIGTRFARQDLEYGASIKKGEWFLIIPSLARDGVFSRDPVDRDGTGASLAYGR 360
DB 301 LRIGTRFARQDLEYGASIKKGEWFLIIPSLARDGVFSRDPVDRDGTGASLAYGR 360
QY 361 GPHVCGVSIARLEAEIAGVTIFRRFPPEMKLKEPVGYPHAPARNIESLNVILKPSKAG 419
DB 361 GPHVCGVSIARLEAEIAGVTIFRRFPPEMKLKEPVGYPHAPARNIESLNVILKPSKAG 419

RESULT 7

US-09-568-472-8
; Sequence 8, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-568-472-8

Query Match 100.0%; Score 2124; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.2e-210;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAFDKFPAGYADPPPAIRLRLEATPIFYWDEGRSVLTRYHDSVA 60
DB 1 MTOEQANQSETKPAFDKFPAGYADPPPAIRLRLEATPIFYWDEGRSVLTRYHDSVA 60
QY 61 VFDERFAVSREMESSAEYSSAIPELSDMKKYGGLPEPDHARVRLKVNPSFTSRAD 120
DB 61 VFDERFAVSREMESSAEYSSAIPELSDMKKYGGLPEPDHARVRLKVNPSFTSRAD 120
QY 121 LRAEIQRITVDQLLDARSGQEEFDVVDYAEIGIPMRASIALKVPACDEKFRFGSATA 180
DB 121 LRAEIQRITVDQLLDARSGQEEFDVVDYAEIGIPMRASIALKVPACDEKFRFGSATA 180
QY 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADSRLS 240
DB 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADSRLS 240
QY 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLRSPALRELVKAEKGRNALDEVLRDNI 300
DB 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLRSPALRELVKAEKGRNALDEVLRDNI 300
QY 301 LRIGTRFARQDLEYGASIKKGEWFLIIPSLARDGVFSRDPVDRDGTGASLAYGR 360
DB 301 LRIGTRFARQDLEYGASIKKGEWFLIIPSLARDGVFSRDPVDRDGTGASLAYGR 360
QY 361 GPHVCGVSIARLEAEIAGVTIFRRFPPEMKLKEPVGYPHAPARNIESLNVILKPSKAG 419
DB 361 GPHVCGVSIARLEAEIAGVTIFRRFPPEMKLKEPVGYPHAPARNIESLNVILKPSKAG 419

RESULT 8

US-09-567-899-8
; Sequence 8, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:

; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/567,899
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 419
; TYPE: PR1
; ORGANISM: Sorangium cellulosum
US-09-567-899-8

Query Match 100.0%; Score 2124; DB 4; Length 419;
Best Local Similarity 100.0%; Pred. No. 1.2e-210;
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTOEQANQSETKPAFDKFPAGYADPPPAIRLRLEATPIFYWDEGRSVLTRYHDSVA 60
DB 1 MTOEQANQSETKPAFDKFPAGYADPPPAIRLRLEATPIFYWDEGRSVLTRYHDSVA 60
QY 61 VFDERFAVSREMESSAEYSSAIPELSDMKKYGGLPEPDHARVRLKVNPSFTSRAD 120
DB 61 VFDERFAVSREMESSAEYSSAIPELSDMKKYGGLPEPDHARVRLKVNPSFTSRAD 120
QY 121 LRAEIQRITVDQLLDARSGQEEFDVVDYAEIGIPMRASIALKVPACDEKFRFGSATA 180
DB 121 LRAEIQRITVDQLLDARSGQEEFDVVDYAEIGIPMRASIALKVPACDEKFRFGSATA 180
QY 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADSRLS 240
DB 181 RALGVGLVPOVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADSRLS 240
QY 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLRSPALRELVKAEKGRNALDEVLRDNI 300
DB 241 TKELVALVGAIIAAGDTTIIYLAFAVLNLRSPALRELVKAEKGRNALDEVLRDNI 300
QY 301 LRIGTRFARQDLEYGASIKKGEWFLIIPSLARDGVFSRDPVDRDGTGASLAYGR 360
DB 301 LRIGTRFARQDLEYGASIKKGEWFLIIPSLARDGVFSRDPVDRDGTGASLAYGR 360
QY 361 GPHVCGVSIARLEAEIAGVTIFRRFPPEMKLKEPVGYPHAPARNIESLNVILKPSKAG 419
DB 361 GPHVCGVSIARLEAEIAGVTIFRRFPPEMKLKEPVGYPHAPARNIESLNVILKPSKAG 419

RESULT 9

US-09-105-537-39
; Sequence 39, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600-438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 416
; TYPE: PR1
; ORGANISM: Streptomyces venezuelae
US-09-105-537-39

```

Query Match      23.0%; Score 488; DB 3; Length 416;
Best Local Similarity 30.5%; Pred. No. 9,4e-42;
Matches 128; Conservative 82; Mismatches 185; Indels 24; Gaps 8;

QY 1 MTOEQANGETPAPDFKFPAGYADDPFPAERLEATPI--FYWDEG-RSWSLTRYHD 57
DB 1 MTRTOGTASPPVLDLGLGQDFADPYPTARLRAEGPAHRVTRPEDEWLVVGYDR 60
QY 58 VSAVFRDERFAVSREWESSAEYSASIPELSDMKKYGFLGFLPPEDHARVKLVNPSFTSR 117
DB 61 ARAVLADPRFS---KQNRNS---TTPLEAEALNNHNMLESPPRTTRKKAWEFTKR 114
QY 118 AIDLRAEIQRTVDQLLDARSQOE--FDVVRDYAEGIPRAISALLKYPACDEKFRFR 175
DB 115 RVELLRPRVQELVDGLVDMMLAPDGRADLMESLAMPPLITYISLGLVPEEDRAAFRW 174
QY 176 GSATRALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLQAEAD 235
DB 175 TDAF-----VFPPDPAQQTAMAMSGYLSRL--IDSKRGQDEEDLSALVTSDD 224
QY 226 GSRLSTKEIVLVAIAGDTTITYLIAFAVLNLRSEPALEIVAAEGKLRNMLDEV 295
DB 225 GSRLTSEELGMVAILLVAGHETTVLVANGVYALSHPDQALALRADMTLLDGAVEEW 284
QY 236 RFDNLRIGTVFAFARODLEYCGASIKKGMVFLIPSLARDGVFSRPPVPRRTGAS 355
DB 235 RREGVESATVFPFEPVLDGTVIPAGDTVLVLADARTBERFPDPHRIORDTAGH 344
QY 336 LAYGPGHYCPGVSLARLEAEIAGVTIFRRFPBMKLETP---VGYPAFPAFNIESLV 411
DB 345 LAFGHGHPICIGAPLARLEARIAVRALRCPDLALDVPGLVWYPMIRIGLKALPI 403

RESULT 10
US-09-252-991A-17836
; Sequence 17836, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17836
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17836

Query Match      22.8%; Score 484.5; DB 4; Length 437;
Best Local Similarity 33.5%; Pred. No. 2.3e-41;
Matches 116; Conservative 74; Mismatches 139; Indels 17; Gaps 5;

QY 50 VWLTRYHVSAYFRDERFAVSREWESSAEYS---SAIPELSDMKKYGFLGFLPPEDHAR 105
DB 60 WVTVTRDARKVLYNH--GVRDARQAALVAKRTSSPRAGSGEGSHHMLNDPPDHNR 117
QY 106 VAKLVNPSFTSRALDLRAEIQRTVDQLLDARSQOEFPVVRDYAEGIPRAISALLKYP 165
DB 118 IASLVGRATTPROVERLQPHIERITBELLDAMAGRCQADLMADFAIPLTIAVIFELGIP 177
QY 166 AECDEKFRFRGSATRALGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRRNPEN 225
DB 178 EAREHARQSWEROALLS-----PEERQALADQVDVLYRL--LEAKRKQPAD-DV 226
QY 226 LTMLOAEADGSLSTKELVALVGAIIAAGDTTITYLIAFAVLNLRSEPALEIVAAEPG 285

```

```

DB 227 YSGLVQADDESGQISEAEIVSAHMLNMSGETTMMIGNLVTLVNPBEOALLRAQOE 286
QY 286 LMRNALDEVLRPNLIRIGTVFAFARODLEYCGASIKKGMVFLIPSLARDGVFSRPPV 345
DB 287 LIPNMEELVRHDSVVRASMLRFTYEDVLDGVTTPAGEIYIVSULTANHDARFDDPDR 346
QY 346 FVVRDGTASLAYGGRPHVCPGVSLARLEAEIAGVTIFRRFPBMKLE 391
DB 347 LDTLNTDGHGIGYGVHYCVGASLARLEGRIALQRLARFPDQL 392

RESULT 11
US-09-320-878-18
; Sequence 18, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: CIP OF 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: CIP OF 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: 60/119,139
; PRIOR FILING DATE: 1999-02-08
; PRIOR APPLICATION NUMBER: 60/100,880
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/087,080
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-320-878-18

Query Match      22.8%; Score 484; DB 3; Length 416;
Best Local Similarity 30.6%; Pred. No. 2.4e-41;
Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQSETPAPDFKFPAGYADDPFPAERLEATPI--FYWDEG-RSWSLTRYHVSAY 61
DB 5 QOQTASPPVLDLGLGQDFADPYPTARLRAEGPAHRVTRPEDEWLVVGYDRARAV 64
QY 62 FRDERFAVSREWESSAEYSASIPELSDMKKYGFLGFLPPEDHARVKLVNPSFTSRALD 121
DB 65 LADPRFS---KQNRNS---TTPLEAEALNNHNMLESPPRTTRKKAWEFTKR 118
QY 122 LRAEIQRTVDQLLDARSQOE--FDVVRDYAEGIPRAISALLKYPACDEKFRFRGSAT 179
DB 119 LRPVQELVDGLVDMMLAPDGRADLMESLAMPPLITYISLGLVPEEDRAAFRWTDFA 178
QY 180 ARAVLGVGLVPOVDEETKTLVASVTEGLALLHDVLDERRRNPENLVLTMLQAEADGSL 239
DB 179 -----VFPPDPAQQTAMAMSGYLSRL--IDSKRGQDEEDLSALVTSDDGSL 228
QY 240 STKELVALVGAIIAAGDTTITYLIAFAVLNLRSEPALEIVAAEGKLRNMLDEVLRPN 299
DB 229 TSEELGMVAILLVAGHETTVLVANGVYALSHPDQALALRADMTLLDGAVEEMLRYEG 288
QY 300 IIRIGTVFAFARODLEYCGASIKKGMVFLIPSLARDGVFSRPPVPRRTGASLAYG 359
DB 289 PVESATVFPFEPVLDGTVIPAGDTVLVLADARTBERFPDPHRIORDTAGHLAEG 348

```

QY 360 RGHVCGVSLARLEAIVGTFRRPEMKLKEP---VFGYHAFRNIESLNV 411
 Db 349 HGHFCTGAPLARLEAIVRALLERCPDLALDVSFGELVWYPMIRGKALPI 403

RESULT 12

US-09-141-908-13
 Sequence 13, Application US/09141908

Patent No. 6503741
 GENERAL INFORMATION:

APPLICANT: ASHLEY, Gary
 APPLICANT: BETLACH, Melanie C.

APPLICANT: BETLACH, Mary
 APPLICANT: MCDANIEL, Robert

APPLICANT: TANG, Li
 TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a

FILE REFERENCE: 30622002100
 CURRENT APPLICATION NUMBER: US/09/141,908

CURRENT FILING DATE: 1998-08-28
 EARLIER APPLICATION NUMBER: CIP OF 09/073,538

EARLIER FILING DATE: 1998-05-06
 EARLIER APPLICATION NUMBER: CIP OF 08/846,247

EARLIER FILING DATE: 1997-04-30
 EARLIER APPLICATION NUMBER: PROV. 60/076,919

EARLIER FILING DATE: 1998-03-05
 EARLIER APPLICATION NUMBER: PROV. 60/087,080

EARLIER FILING DATE: 1998-05-28
 NUMBER OF SEQ ID NOS: 31

SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 13

LENGTH: 416
 TYPE: PR

ORGANISM: Streptomyces venezuelae
 US-09-141-908-13

Query Match 22.8%; Score 484; DB 4; Length 416;
 Best Local Similarity 30.6%; Pred. No. 2,4e-41;

Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQSEKPAFPKPPAPGYADPPFAIERLEATPI--FYWDEG-RSWVLTTRYHDSAV 61
 Db 5 QCGTTASPPVLIDGALGQDPADPYTYARLRABEGPAHVRTEPEGDEVLVGCDARAV 64
 QY 62 FRDERPAVREEMESSAEYSASAIPELSDMKYGLFGLPREDHARVRLVNPSTSRIDL 121
 Db 65 LADPRS---KDMRNS---TTPLTEAEALNHNMLSDPPRHTRLRLKLVAREFTMRVEL 118
 QY 122 LRAEIORTVDQLDARSQGE--FDVVRDYAEGIPRAISALLKVAECDEKFRFGSAT 179
 Db 119 LRPRVOEIVDGLVADNMAAPDGRADLMSIAMPPLPTIVISELGVPEPRAARVWDAF 178
 QY 180 ARAAGVGLPQVDEEKTIVASVTEGIALHDLVDERRRNPLENDVLTMLQAEADGSR 239
 Db 179 -----VFPDDPAQQTAAEMSGYLSRL---IDSKRGQDGEDLSALVRTSDEGSRL 228
 QY 240 STEELVALGAIITAACTDTTITVIAFAVNLNLSPEALVYAEPCGMNADDEVIRFDN 299
 Db 229 TSEELGMAHILLVAGHETTIVNLIANGVYALLSHPDQALARDMTLLDGAVEEMLRYEG 288
 QY 300 ILRIGTRFARQDLEYGASIKKGEVFLIISALRDGTVFSRPDVFVRDGTGASLAVG 359
 Db 289 PVSATYRFEVEVDLDGTVIPAGDVLVVLADAHRTPERFPDRDIRDRTAGHLAFG 348
 QY 360 RGHVCGVSLARLEAIVGTFRRPEMKLKEP---VFGYHAFRNIESLNV 411
 Db 349 HGHFCTGAPLARLEAIVRALLERCPDLALDVSFGELVWYPMIRGKALPI 403

RESULT 13
 US-09-657-440-18
 Sequence 18, Application US/09657440

Patent No. 6509455
 GENERAL INFORMATION:
 APPLICANT: ASHLEY, Gary
 APPLICANT: BETLACH, Melanie C.

APPLICANT: BETLACH, Mary C.
 APPLICANT: MCDANIEL, Robert

APPLICANT: TANG, Li
 TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

FILE REFERENCE: 30622002120
 CURRENT APPLICATION NUMBER: US/09/657,440

CURRENT FILING DATE: 2000-09-07
 PRIOR APPLICATION NUMBER: 09/320,878

PRIOR FILING DATE: 1999-05-27
 PRIOR APPLICATION NUMBER: CIP OF 09/141,908

PRIOR FILING DATE: 1998-08-28
 NUMBER OF SEQ ID NOS: 34

SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 18

LENGTH: 416
 TYPE: PR

ORGANISM: Streptomyces venezuelae
 US-09-657-440-18

Query Match 22.8%; Score 484; DB 4; Length 416;
 Best Local Similarity 30.6%; Pred. No. 2,4e-41;

Matches 127; Conservative 81; Mismatches 183; Indels 24; Gaps 8;

QY 5 QANQSEKPAFPKPPAPGYADPPFAIERLEATPI--FYWDEG-RSWVLTTRYHDSAV 61
 Db 5 QCGTTASPPVLIDGALGQDPADPYTYARLRABEGPAHVRTEPEGDEVLVGCDARAV 64
 QY 62 FRDERPAVREEMESSAEYSASAIPELSDMKYGLFGLPREDHARVRLVNPSTSRIDL 121
 Db 65 LADPRS---KDMRNS---TTPLTEAEALNHNMLSDPPRHTRLRLKLVAREFTMRVEL 118
 QY 122 LRAEIORTVDQLDARSQGE--FDVVRDYAEGIPRAISALLKVAECDEKFRFGSAT 179
 Db 119 LRPRVOEIVDGLVADNMAAPDGRADLMSIAMPPLPTIVISELGVPEPRAARVWDAF 178
 QY 180 ARAAGVGLPQVDEEKTIVASVTEGIALHDLVDERRRNPLENDVLTMLQAEADGSR 239
 Db 179 -----VFPDDPAQQTAAEMSGYLSRL---IDSKRGQDGEDLSALVRTSDEGSRL 228
 QY 240 STEELVALGAIITAACTDTTITVIAFAVNLNLSPEALVYAEPCGMNADDEVIRFDN 299
 Db 229 TSEELGMAHILLVAGHETTIVNLIANGVYALLSHPDQALARDMTLLDGAVEEMLRYEG 288
 QY 300 ILRIGTRFARQDLEYGASIKKGEVFLIISALRDGTVFSRPDVFVRDGTGASLAVG 359
 Db 289 PVSATYRFEVEVDLDGTVIPAGDVLVVLADAHRTPERFPDRDIRDRTAGHLAFG 348
 QY 360 RGHVCGVSLARLEAIVGTFRRPEMKLKEP---VFGYHAFRNIESLNV 411
 Db 349 HGHFCTGAPLARLEAIVRALLERCPDLALDVSFGELVWYPMIRGKALPI 403

RESULT 14

5212296-9

Patent No. 5212296
 APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH

J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
 TEPPERMAN, JAMES M.

TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
 CYTOCHROMES

NUMBER OF SEQUENCES: 19
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/569,781
 FILING DATE: 23-AUG-1990

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 464,499

FILING DATE: 12-JAN-1990
 APPLICATION NUMBER: 405,605

FILING DATE: 11-SEP-1989

SEQ ID NO:9:
LENGTH: 403
5212296-9

Query Match 21.6%; Score 459; DB 6; Length 403;
Best Local Similarity 31.7%; Pred. No. 8.8e-39;
Matches 127; Conservative 73; Mismatches 157; Indels 44; Gaps 12;

QY 21 ARGVADDEPPAERLEATPI---FYWDEGRSVLTRYHVSADVDEREFASREWESS 77
DB 20 AFGCFDDAPDVTEAARTEPVTRATLMDGSCVLTRODVRVAGDPF-----SA 71
QY 78 AEYSAPISLSDMKYKFGFLP-----EDHAFVKLVNPSFTSRAIDLRAEIORTV 130
DB 72 DAHRTGFPLTAGR-EIIGTNPFLRMDPEHARLRMLTDPIVKYKEARREPVQRLA 130
QY 131 DQOLD-AASGGEFVVDYAGIEMRAISALKTPACDEKFRFGSATARALGVGLVP 189
DB 131 DDLVDMRTTGRSADLVTEFALPLPSLVICLLGVPEY-DHAF--FOERSVLTLRSTP 187
QY 190 QVDEETKTLVASVTEGLALHDVLDERRRNPLENDVLTMLQAEADGSRSLSTKELVALVG 249
DB 188 ---EVRAAQDELELYLARL---ARTKREPPDAIISRUVARGELDDQIATMGRLLV- 240
QY 250 AIIAGDTTTYLLAFVAVLNLRSPEALELYAEKRLNALDEVLRPDNILRIGTVFA 309
DB 241 ---AGHETANMTALSTVLRLNPDQLARLAEPALVGAVEELLRYLTIHNGVPRIA 296
QY 310 RODLEYCGASIKKGMVFLILPSALRDGTVSRPDVFRDPTGASLAYGCPHYCPGV 369
DB 297 TEDVIGRTTIAAGGVCMISSANRDAEVPFGDLDVARDARRHVAFGFVHQCLGP 356
QY 370 LARLEAEIAGTIFRRPEMKL---KETP-----VFQYH 400
DB 357 LARVELQIAIETLRRLPDRLAVPHEIIFRGMATYGVH 397

RESULT 15
US-09-252-991A-32437
; Sequence 32437, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32437
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32437

Query Match 19.8%; Score 421.5; DB 4; Length 468;
Best Local Similarity 29.5%; Pred. No. 8.3e-35;
Matches 133; Conservative 70; Mismatches 177; Indels 71; Gaps 15;

QY 5 QANQSETPA-----FDFKPFAGYADEPPAERLEATPIFYWDEGRSVLTRYHD 57
DB 36 QPRHDAQRPALAPRSDGPDHTYHPDFAVDPPLRLIRSAVCR-DQASIMWISRYAD 94
QY 58 VSAVFRDRAFAVS-----REEMSSAEYSASAIPELSDMKYKGLFCLPPEDHARVRK 108
DB 95 VSACURDRRFSADPARLGAAGVRCGASWFGHQQLPLARFYDNFMLENDAPR-HTRLRK 153
QY 109 LVNPSFTSRAIDLRAEIORTVQDILLARSQGEFVDVVDYAGIEMRAISALKVPAEC 168

DB 154 LPAPAFGPDVARRWEARIEVLVEELDSLERRRBDLLRDAEPITTVAAELFGFPRED 213
QY 169 DEKFRPFGSATARALGVLPQVDEETLVAS-----VTEGLALHDVLDERRR- 219
DB 214 TGQLLPWG---RDLAAGL-----DLASHGDACQINRSAAAFSDYLQROARKMSD 260
QY 220 -----PLENDVLTMLQAEADGSRSLSTKELVALVGAIIAAGTDTTYLLAFVAVLN 269
DB 261 GSNRPFGSAASIFDGAAML---EAG---LGLLEDVAAVAVFMAAFETITSMGNATLA 314
QY 270 LIRSPEALELYKAEBGLRNALDEVLRPDNILRIGTVFARQDLEYCGASIKKGMVFL 329
DB 315 LITHPDQDLIRRCPEILANAVVEELIRPDGAVR-GGVACTLEEVEIGGQRIIPGGEKWL 373
QY 330 IPSALRDGTVSRPDVDPVR-DTGASLAYGCPHYCGVSLARLEAEIIV-GTIFRRF- 386
DB 374 FLAANRDEMPAAPRLOLQORANAKOHVAFHGPYICGAYLARLELOCALGLVRRFA 433
QY 387 -----PEMKKETPVFGYHPAFRNIESLNV 412
DB 434 LASEPTDLRWRRSV-----FRTLERLPV 458

Search completed: October 2, 2003, 17:35:54
Job time: 12.2473 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:27:29 ; Search time 28.237 Seconds

(without alignments)
2355.298 Million cell updates/sec

Title: US-10-014-717-8

Perfect score: 2124
Sequence: 1 MTOEQANQSEFKPAFDKPF.....HPAFNIRSLNVLKPSKAG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneeq/geneeqp-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2124	100.0	419	21	AA198579
2	528.5	24.8	396	21	AA198502
3	528.5	24.7	396	21	AA198503
4	518.5	24.4	396	21	AA198501
5	516	24.3	399	24	ABP57702
6	491	23.1	475	24	AAE34729
7	488.5	23.0	475	24	AAE34732
8	486.5	22.9	470	24	AAE34724
9	485.5	22.9	430	24	AAE34736

10	484	22.8	416	21	AA198654
11	484	22.8	416	21	AA198728
12	484	22.8	416	24	ABG71678
13	480.5	22.6	430	24	AAE34780
14	476.5	22.4	415	21	AA197196
15	475.5	22.4	430	24	AAE34725
16	475.5	22.4	430	24	AAE34727
17	475.5	22.4	430	24	AAE34731
18	466	21.9	432	24	AAE34726
19	463.5	21.8	430	24	AAE34737
20	452.5	21.3	430	24	AAE34730
21	452	21.3	482	23	ABB81316
22	451.5	21.3	429	24	AAE34728
23	450	21.2	396	23	ABB81315
24	449	21.1	448	24	ABU11985
25	448	21.1	430	24	AAE34723
26	448	21.1	430	24	AAE34735
27	447	21.0	430	24	AAE34733
28	443.5	20.9	430	24	AAE34734
29	443	20.9	399	24	ABU11376
30	435.5	20.5	426	24	AAE34738
31	435.5	20.5	430	24	AAE34722
32	432	20.3	403	12	AA111350
33	420	19.8	411	19	AAW54389
34	414.5	19.5	404	12	AA14724
35	413	19.4	396	15	AAE60777
36	407.5	19.2	587	18	AAW33274
37	406.5	19.1	407	24	ABP95990
38	405.5	19.1	418	21	AB13589
39	403.5	19.0	420	21	AA193790
40	401.5	18.9	402	24	ABU11981
41	386	18.2	398	22	AA131560
42	385.5	18.1	405	24	ABP95987
43	382	18.0	406	24	ABP95992
44	371	17.5	409	22	AAE07915
45	370.5	17.4	398	17	AA11585

ALIGNMENTS

RESULT 1	
AA198579	standard; Protein, 419 AA.
AA198579;	
AC	AA198579;
XX	
DT	10-APR-2000 (first entry)
XX	
DE	Sorangium cellulosum epothilone macrolactone oxidase EPOS F.
XX	
KM	EPOS F, epothilone macrolactone oxidase; epothilone biosynthesis; taxol substitute; anticancer.
XX	
OS	Sorangium cellulosum.
XX	
PM	MO9966028-A2.
PD	23-DEC-1999.
XX	
PF	16-JUN-1999; 99WO-EP04171.
XX	
PR	18-JUN-1998; 98US-0099504.
XX	
PA	24-SEP-1998; 98US-0101631.
XX	
XX	05-FEB-1999; 99US-0118906.
PA	(NOVS) NOVARTIS AG.
XX	(NOVS) NOVARTIS-ERRINDUNGEN VERN GRS MBH.
PI	Schupp T, Ligon JM, Molnar I, Zirkle R, Geerlach J, Cyr D;
XX	WPI; 2000-097747/08.
DR	N-PSDB; AA255887.

Amino acid sequenc
Macrolide ring mod
S. venezuelae plic
Streptomycetes tuber
S. venezuelae macr
Streptomycetes lydic
Streptomycetes chaib
Streptomycetes kasug
Streptomycetes speci
Streptomycetes lydic
Streptomycetes plate
Streptomycetes sp. c
Streptomycetes speci
Streptomycetes sp. c
M. echinospora cal
Streptomycetes tuber
Streptomycetes plate
Streptomycetes tuber
Streptomycetes plate
Protein encoded by
Streptomycetes tuber
Streptomycetes tuber
Cytochrome enzyme
Actinomadura hibis
6-hydroxylase enco
Mycinnamycin IV hyd
S. fradiae tylosin
Microtetraospora re
Streptomycetes globi
S. fradiae tylosin
M. echinospora cal
Pimaricin biosynth
Streptomycetes sp. T
Streptomycetes lydic
S. clavuligerus pr
Streptomycetes prist

AAB15503
ID AAB15503 standard; Protein; 396 AA.
XX
AC AAB15503;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating protein #3.
XX
KM Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
XX hydroxymethylglutaryl co-enzyme A reductase; hypcholesterolaemia.
OS Bacillus subtilis.
XX WO200044886-A1.
XX PN 03-AUG-2000.
XX PD
XX 28-JAN-2000; 2000WO-JP00472.
XX PF
XX PR 29-JAN-1999; 99JP-0021707.
XX (KYO) KYOMA HAKKO KOGYO KK.
XX PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX DR WPI: 2000-548827/50.
XX N-PSDB; AAA95669.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors -
PS Claim 10; Page 105-107; 111pp; Japanese.
XX
CC This sequence represents a novel Bacillus derived protein having
CC the activity of producing a hydroxylated bicyclic compound or the
CC corresponding lactone from a bicyclic compound or the corresponding
CC lactone. The protein is used for preparing hydroxylated compounds
CC useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC inhibitors and hypcholesterolaemic agents.
CC
SQ Sequence 396 AA;
Query Match 24.7%; Score 525.5; DB 21; Length 396;
Best Local Similarity 33.2%; Pred. No. 3.2e-42; Indels 29; Gaps 6;
Matches 126; Conservative 71; Mismatches 154;
QY 25 AEDPPAIERLREATPIFYWDEGRSVLTRYHDVSAVFRDERFAVSREEMESSAEYSSAI 84
DB 23 AVHPFWYESMRKDAVPSFDEENQWVSFLYDDVKVGDKEI-----FSSCM 70
QY 85 PELSMKKYGLFGLPEDHARVKLVNPSFTSRADLRAEIQRTVDLDRSGOEED 144
DB 71 POOTSSIGNSIISMPPKTKIRSVNKAFTPRAMQWEPRIQETIDELIQFGSRSEPD 130
QY 145 VVRDVAEGIPMRASIALTKVPAECDEKFRFGSATARALGVLPQVDEETTLVA---- 200
DB 131 LVHDSYLPVIVISLGVPSAHNEQFKMSD-----LVSTPKDKSEAEAKFLSEED 185
QY 201 SVTEGIALLDVLDERRNPLENDVLTMLQAEADGSRSLSTKELVALVGAIITAGDTTIT 260
DB 186 KCEELAAFAFGIIEKKNKPEODIISILVEAETGKSGELIPFCTLLLVAGNETTT 245
QY 261 YLIAAVNLRLSPALVKAEPGLMRNALDEVLRF---DNILRIGTVRPARODLEYCG 317
DB 246 NLIISAMYSILTEPGVIELRSHBELMPQAVEALRFAPAPVLR-----RIAKRDTETIGG 301
QY 318 ASIKKGEMVFLIPSLRDPVFSRPDVEDVDRDQASLAYSGRGPHVCPGVSILARLEARI 377
DB 302 HLIKEGDVTYLAVASANRDEAKFDRPHMDIRHNPPIAFGHGIFHCGLAPLARLEANI 361
QY 378 AVGTIFRRPEMK-LKETPV 396

DB 362 ALTSLISAFPHMECVSITPI 381
RESULT 4
ID AAB15501 standard; Protein; 396 AA.
XX
AC AAB15501;
XX
DT 14-FEB-2001 (first entry)
XX
DE Bacillus subtilis hydroxylating protein.
XX
KM Hydroxylated bicyclic compound; lactone; HMG-CoA; inhibitor;
XX hydroxymethylglutaryl co-enzyme A reductase; hypcholesterolaemia.
OS Bacillus subtilis.
XX WO200044886-A1.
XX PN 03-AUG-2000.
XX PD
XX 28-JAN-2000; 2000WO-JP00472.
XX PF
XX PR 29-JAN-1999; 99JP-0021707.
XX (KYO) KYOMA HAKKO KOGYO KK.
XX PI Endo H, Yonetani Y, Mizoguchi H, Hashimoto S, Ozaki A;
XX DR WPI: 2000-548827/50.
XX N-PSDB; AAA95669.
XX
PT New protein derived from Bacillus genus microorganism useful for
PT producing a hydroxylated bicyclic compound e.g. HMG-CoA reductase
PT inhibitors -
PS Claim 8; Page 68-71; 111pp; Japanese.
XX
CC This sequence represents a novel Bacillus derived protein having
CC the activity of producing a hydroxylated bicyclic compound or the
CC corresponding lactone from a bicyclic compound or the corresponding
CC lactone. The protein is used for preparing hydroxylated compounds
CC useful as hydroxymethylglutaryl co-enzyme A (HMG-CoA) reductase
CC inhibitors and hypcholesterolaemic agents.
CC
SQ Sequence 396 AA;
Query Match 24.4%; Score 518.5; DB 21; Length 396;
Best Local Similarity 32.6%; Pred. No. 1.5e-41; Indels 29; Gaps 6;
Matches 124; Conservative 72; Mismatches 155;
QY 25 AEDPPAIERLREATPIFYWDEGRSVLTRYHDVSAVFRDERFAVSREEMESSAEYSSAI 84
DB 23 AVHPFWYESMRKDAVPSFDEENQWVSFLYDDVKVGDKEI-----FSSCM 70
QY 85 PELSMKKYGLFGLPEDHARVKLVNPSFTSRADLRAEIQRTVDLDRSGOEED 144
DB 71 POOTSSIGNSIISMPPKTKIRSVNKAFTPRAMQWEPRIQETIDELIQFGSRSEPD 130
QY 145 VVRDVAEGIPMRASIALTKVPAECDEKFRFGSATARALGVLPQVDEETTLVA---- 200
DB 131 LVHDSYLPVIVISLGVPSAHNEQFKMSD-----LVSTPKDKSEAEAKFLSEED 185
QY 201 SVTEGIALLDVLDERRNPLENDVLTMLQAEADGSRSLSTKELVALVGAIITAGDTTIT 260
DB 186 KCEELAAFAFGIIEKKNKPEODIISILVEAETGKSGELIPFCTLLLVAGNETTT 245
QY 261 YLIAAVNLRLSPALVKAEPGLMRNALDEVLRF---DNILRIGTVRPARODLEYCG 317
DB 246 NLIISAMYSILTEPGVIELRSHBELMPQAVEALRFAPAPVLR-----RIAKRDTETIGG 301

Oy	48	RSWLTETVHDSVAFDEREPAVSRREWESAEYSNAIPELS-----DKKXGGLGP---	93
Db	94	PIMLVTRFDVAREVMQRQF-----VNNPFLVPEIGADDDPRARLIELFGIPEDL	1433
Oy	100	-----PEDHARVAKLVNPSFTSRAIDLRAEIQRTVDQLDARSGOEE---FD	1444
Db	144	APVLTDTILNSDDPPDHRLRLRVLSRAFTARRIQDLRPRERTTDELLARLPDHAEDEV	2013
Oy	145	VVRDYAEGIMRAISALKVPAECDEKEFRFGSATSARALGVLPQVDEFTKLVASYTE	2044
Db	204	LVEHFAPVLEPIITYICEVIGVDEEDRALMRFFGADLA-----SLNP-----KRIGATWPE	2523
Oy	205	GLALHVLVBERRNPLNDVLMMLQA-EADGSRSTKEVLAIVGATIAAGDTPTTYLI	2633
Db	253	MIHIVHIDERRAD-LRDDLSGLLRADDDGGRISDVEMVTLVTLVGHETTHALI	3111
Oy	264	AFAYLVNLSRPEALEVKAEPGLMRNALDEVLFNDILRICTVRFARODLEYGASIKKG	3223
Db	312	SNGLTALVTHPDQRRLIDEDPALLPRAVHELMRWGCPIGATQLRARMEDTEVAGVQRQG	3711
Oy	324	EMVFLILPSLRKGVTPSRPDVVDVRRDTGA-----SLAYGRGPHVCPGVSLARLEAETAV	3793
Db	372	EALMFSTLVAAHHDPRRHYTGPERLDLTRQAGRAEDHVGFGHGMHYCLGASLARQAEVAY	4311
Oy	380	GTIFRPEPKLKETP	395
Db	432	GKLTIRYFDLALALTP	447
RESULT 7			
ID	AAE34732	standard; Protein; 475 AA.	
AC	AAE34732:		
XX	14-MAY-2003	(first entry)	
DE	Streptomyces rimosus emali protein.		
KM	P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;		
KW	emamectin; insecticide; emali protein.		
XX	Streptomyces rimosus.		
OS	WO200292801-A2.		
PN	21-NOV-2002.		
XX	15-MAY-2002; 2002MO-EP05363.		
PF	16-MAY-2001; 2001US-291149P.		
XX	(SYGN) SYNGENTA PARTICIPATIONS AG.		
XX	Mojnar I, Ligon JW, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;		
PI	Buckel TG;		
XX	WPI; 2003-140280/13.		
DR	N-PSDB; AAD53027.		
XX	Novel polypeptide that exhibits enzymatic activity of P450		
PT	monooxygenase and capable of regioselectively oxidizing alcohol of		
PT	avermectin useful for making emamectin from avermectin		
XX	Claim 17, Page 121-122; 157pp; English.		
XX	The present invention relates to novel proteins that exhibit an enzymatic		
CC	activity of P450 monooxygenase and capable of regioselectively oxidizing		
CC	the 4-cadinol group at position 4 of avermectin to 4-keto-avermectin.		
CC	The invention also relates to ferrodoxins and ferrodoxin reductases that		
CC	are active with the P450 monooxygenases. Sequences of the invention are		
CC	useful for producing 4-keto-avermectin from avermectin, which is useful		

CC for producing emamectin. Emamectin is useful as an insecticide. The
CC present sequence is *Streptomyces rimosus* emali protein.

Sequence 475 AA;

Query Match	Score	DB	Length
Best Local Similarity.	23.0%;	488.5;	24;
	29.7%;	Pred. No. 1.7e-38;	

Matches 131; Conservative 83; Mismatches 158; Indels 69; Gaps 12;

QY 2 TQEQANQSETK-PAFDFKPFAPGYA-----EDPFAIERLREATPIF--- 42

Db 29 TPSAATSSDTTYPATTDRITLPSYVGLHPGEPNLMEPELDNPTGYGTLLREQAPLVRAR 88

43 YWDEGRSWLTRYHDVSAVFRDERFAVSREEMESSAEYSSAIPELSDMKK-----YGLFG 97

Db 89 FIDDSPIWLTRFDVREVRDQRF-----VNNPTLVPGIGADKDPARLIEFG 138

QY 98 LP-----PEDHARVRKLVNPSFTSRAIDLRAEIQRTVDQLDARSGQE 142

Db 139 IPEDLTPYLADTILTSDDPPDHTRLRLVSRFAFTARRIQDLRPRVEQITDALLERLPDHA 198

QY 143 ---FDVVRDYAEGIPMRAISALLKVPACEDKFRFRGSAATARALGVGLVPQVDEETKTLV 199

Db 199 DGVVDLVEHFAYPLPTVICELVGIDEEDRTLWRRFGADLA-----SLNP-----KRIG 247

200 ASVTGIALHDLVDERBRNPLENDVLTMLLOA-EADGSRISTKEI VALVGAITAACTDT 258

248 ATMPENTAHTEYIDERRA-ALPDDI.SGLTPADDDGRI.SDVEMYTI.VI.TI.VI.AGHET 306

259 TIVYIAEAVTNIIBSEPAIFLYKAEBCIMBNALDEVIPEUNNITIPICTVBFAPODIEVCGA 318

[illegible][illegible][illegible]

1.
 2.
 3.
 4.
 5.
 6.
 7.
 8.
 9.
 10.
 11.
 12.
 13.
 14.
 15.
 16.
 17.
 18.
 19.
 20.
 21.
 22.
 23.
 24.
 25.
 26.
 27.
 28.
 29.
 30.
 31.
 32.
 33.
 34.
 35.
 36.
 37.
 38.
 39.
 40.
 41.
 42.
 43.
 44.
 45.
 46.
 47.
 48.
 49.
 50.
 51.
 52.
 53.
 54.
 55.
 56.
 57.
 58.
 59.
 60.
 61.
 62.
 63.
 64.
 65.
 66.
 67.
 68.
 69.
 70.
 71.
 72.
 73.
 74.
 75.
 76.
 77.
 78.
 79.
 80.
 81.
 82.
 83.
 84.
 85.
 86.
 87.
 88.
 89.
 90.
 91.
 92.
 93.
 94.
 95.
 96.
 97.
 98.
 99.
 100.

RESULT 8

ID AAE34724 standard;. Protein; 470 AA.

AAE34724;

DT 14-MAY-2003 (first entry)
XX

Streptomyces rimosus emaz protein

XX
XX
B450 m0n0c0v0c0n0s0s0: f0r0r0d0v0i0n: f0r0d0v0i0n r0d0i0c0s0s0: 0n0z0m0:

KW emamectin; insecticide; emaz protein.
 YX

Streptomyces rimosus.

PN WO200292801-A2.

PD 21-NOV-2002.

PF 15-MAY-2002; 2002WO-EP05363.

PR 16-MAY-2001; 2001US-291149P.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;

XX 1
1 2
2 3
3 4
4 5
5 6
6 7
7 8
8 9
9 10
10 11
11 12
12 13
13 14
14 15
15 16
16 17
17 18
18 19
19 20
20 21
21 22
22 23
23 24
24 25
25 26
26 27
27 28
28 29
29 30
30 31
31 32
32 33
33 34
34 35
35 36
36 37
37 38
38 39
39 40
40 41
41 42
42 43
43 44
44 45
45 46
46 47
47 48
48 49
49 50
50 51
51 52
52 53
53 54
54 55
55 56
56 57
57 58
58 59
59 60
60 61
61 62
62 63
63 64
64 65
65 66
66 67
67 68
68 69
69 70
70 71
71 72
72 73
73 74
74 75
75 76
76 77
77 78
78 79
79 80
80 81
81 82
82 83
83 84
84 85
85 86
86 87
87 88
88 89
89 90
90 91
91 92
92 93
93 94
94 95
95 96
96 97
97 98
98 99
99 100
100 101
101 102
102 103
103 104
104 105
105 106
106 107
107 108
108 109
109 110
110 111
111 112
112 113
113 114
114 115
115 116
116 117
117 118
118 119
119 120
120 121
121 122
122 123
123 124
124 125
125 126
126 127
127 128
128 129
129 130
130 131
131 132
132 133
133 134
134 135
135 136
136 137
137 138
138 139
139 140
140 141
141 142
142 143
143 144
144 145
145 146
146 147
147 148
148 149
149 150
150 151
151 152
152 153
153 154
154 155
155 156
156 157
157 158
158 159
159 160
160 161
161 162
162 163
163 164
164 165
165 166
166 167
167 168
168 169
169 170
170 171
171 172
172 173
173 174
174 175
175 176
176 177
177 178
178 179
179 180
180 181
181 182
182 183
183 184
184 185
185 186
186 187
187 188
188 189
189 190
190 191
191 192
192 193
193 194
194 195
195 196
196 197
197 198
198 199
199 200
200 201
201 202
202 203
203 204
204 205
205 206
206 207
207 208
208 209
209 210
210 211
211 212
212 213
213 214
214 215
215 216
216 217
217 218
218 219
219 220
220 221
221 222
222 223
223 224
224 225
225 226
226 227
227 228
228 229
229 230
230 231
231 232
232 233
233 234
234 235
235 236
236 237
237 238
238 239
239 240
240 241
241 242
242 243
243 244
244 245
245 246
246 247
247 248
248 249
249 250
250 251
251 252
252 253
253 254
254 255
255 256
256 257
257 258
258 259
259 260
260 261
261 262
262 263
263 264
264 265
265 266
266 267
267 268
268 269
269 270
270 271
271 272
272 273
273 274
274 275
275 276
276 277
277 278
278 279
279 280
280 281
281 282
282 283
283 284
284 285
285 286
286 287
287 288
288 289
289 290
290 291
291 292
292 293
293 294
294 295
295 296
296 297
297 298
298 299
299 300
300 301
301 302
302 303
303 304
304 305
305 306
306 307
307 308
308 309
309 310
310 311
311 312
312 313
313 314
314 315
315 316
316 317
317 318
318 319
319 320
320 321
321 322
322 323
323 324
324 325
325 326
326 327
327 328
328 329
329 330
330 331
331 332
332 333
333 334
334 335
335 336
336 337
337 338
338 339
339 340
340 341
341 342
342 343
343 344
344 345
345 346
346 347
347 348
348 349
349 350
350 351
351 352
352 353
353 354
354 355
355 356
356 357
357 358
358 359
359 360
360 361
361 362
362 363
363 364
364 365
365 366
366 367
367 368
368 369
369 370
370 371
371 372
372 373
373 374
374 375
375 376
376 377
377 378
378 379
379 380
380 381
381 382
382 383
383 384
384 385
385 386
386 387
387 388
388 389
389 390
390 391
391 392
392 393
393 394
394 395
395 396
396 397
397 398
398 399
399 400
400 401
401 402
402 403
403 404
404 405
405 406
406 407
407 408
408 409
409 410
410 411
411 412
412 413
413 414
414 415
415 416
416 417
417 418
418 419
419 420
420 421
421 422
422 423
423 424
424 425
425 426
426 427
427 428
428 429
429 430
430 431
431 432
432 433
433 434
434 435
435 436
436 437
437 438
438 439
439 440
440 441
441 442
442 443
443 444
444 445
445 446
446 447
447 448
448 449
449 450
450 451
451 452
452 453
453 454
454 455
455 456
456 457
457 458
458 459
459 460
460 461
461 462
462 463
463 464
464 465
465 466
466 467
467 468
468 469
469 470
470 471
471 472
472 473
473 474
474 475
475 476
476 477
477 478
478 479
4

DR WEI, 2003-140280/13;
N-PSDB: AAD53019.
DR

PT Novel polypeptide that exhibits enzymatic activity of P450
 PT monooxygenase and capable of regioselectively oxidizing alcohol of
 avermectin useful for making emamectin from avermectin -
 PS Claim 17; Page 107-108; 157pp; English.

CC The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidizing
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferredoxins and ferredoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces rimosus emas protein.

XX Sequence 470 AA;

Query Match 22.9%; Score 486.5; DB 24; Length 470;
 Best Local Similarity 30.6%; Pred. No. 2.6e-38;
 Matches 124; Conservative 81; Mismatches 147; Indels 53; Gaps 10;

QY 22 PGAAEDPPAIEKRLREAPF--YWDGRSVLTRYHDVSAVFRDERAVSRREWESSA 78
 DB 60 PELLEPPYGYGTLRQAPLVRRARFIDDSPIMLVTRFDVVRVREVRDQRF-----V 109

QY 79 EYSSAIPBELSDMKK---YGLFGLP-----PDHARVRLVNPSTSR 118
 DB 110 NNPTLVPGICADDPARARLIEFGIPEDLAPYLTDNLTSDPDHTRRLVRAFAFR 169

QY 119 IDLLRAEIQRTVDQLDARSQEE--FDVVRDYAGIPIRAISALLKVPACDEKFRFP 175
 DB 170 IQLRRVERITDELRLRPHADGVVDLVEHFAVPLPTVI CELVGI DEEDRALMRFP 229

QY 176 GSATAALGVLPQVDEETKTIVASTEGALLHVDLDRRRNPENDVLTMLQA-EA 224
 DB 230 GAOLA-----SLNP-----KRIGATMPEMISHIHETIDERRA-ALRDDLGLIRQDD 277

QY 235 DGSRLSTKELVALVGAIIAAGTDTTYLAFAVLNLSPEALELVKAEPLGRNMLDEV 294
 DB 278 DGSRLDEVNTLVTLVLAAGHETTAHLSNGTALLTHDQRLDIDEDALPRAVHEL 337

QY 295 LRFDNLTIGTVAFARODLEYCGASIKGEMVFLIPSAIRDGVSRPDPVRRDTGA 354
 DB 338 MRWCGPIQATQLRYALDETVAGVQVQGEALMFSIVAANHDPRHYTPERLDTIQPAG 397

QY 355 ----SLAYGRPHVCPGVSIARLEAEIANGVTIFRRPEMKLXETP 395
 DB 398 RABDHVGFHGMHYCLGASLARQEADEVAYGKLITRYPDALALTP 442

RESULT 9
 AAE34736
 ID AAE34736 standard; Protein; 430 AA.

XX AAE34736;
 AC
 XX
 DT 14-MAY-2003 (first entry)

DE Streptomyces lydicus emas5 protein.
 XX P450 monooxygenase; avermectin; ferredoxin; ferredoxin reductase; enzyme;
 KM emamectin; insecticide; emas5 protein.

OS Streptomyces lydicus.
 XX
 PN WO200292801-A2.

PD 21-NOV-2002.

PF 15-MAY-2002; 2002MO-EP05363.
 XX
 PR 16-MAY-2001; 2001US-291149P.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;

DR WPI; 2003-140280/13.
 DR N-PSDB; AAD530311.

PT Novel polypeptide that exhibits enzymatic activity of P450
 PT monooxygenase and capable of regioselectively oxidizing alcohol of
 avermectin useful for making emamectin from avermectin -
 PS Claim 17; Page 128-129; 157pp; English.

CC The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidizing
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferredoxins and ferredoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces lydicus emas5 protein.

XX Sequence 430 AA;

Query Match 22.9%; Score 485.5; DB 24; Length 430;
 Best Local Similarity 31.4%; Pred. No. 2.9e-38;
 Matches 129; Conservative 81; Mismatches 150; Indels 51; Gaps 13;

QY 22 PGAAEDPPAIEKRLREAPF--YWDGRSVLTRYHDVSAVFRDERAVSRREWESSA 78
 DB 25 PALIGDPFAGYALREOGFVVRGRFVDDSPVWFVTRFEVREVRDQRF--RNPVSSA 81

QY 79 EYSSAIPBELSDMKK-YGLFGLP-----PDHARVRLVNPSTSR 122
 DB 82 --PDADPEDTPLSRLMDMKGFPEHLRVYLSI LNDADPDHTRRLVRAFAFR 139

QY 123 RAEIQRTVDQLDARSQEE--FDVVRDYAGIPIRAISALLKVPACDEKFRFGSAT 179
 DB 140 RRVQAQIABELLARLPEHNDGVVDLQHFAPLPPTVI CELVGI PEEDRPQWRMGADL 199

QY 180 APALGVLPQVDEETKTIVASTEGALLHVDLDRRRNPENDVLTMLQA-EADGR 238
 DB 200 ----VSLQP--DRMSRSPAMIDH----IHEIIAARR-ALTDLISLITRTHDDGGR 247

QY 239 LSTKEIHALVGAIIAAGTDTTYLAFAVLNLSPEALELVKAEPLGRNMLDEVLRD 298
 DB 248 LSDVEVNTLVTLVLAAGHETTAHLSNGTALLTHDQRLDIDEDALPRAVHELKRC 307

QY 299 NILRIGTVAFARODLEYCGASIKGEMVFLIPSAIRDGVSRPDPVRRDTGA---- 354
 DB 308 GPVHMQQLRYADEVGLAVGRIRKGDVQGLIVSANRDRHYTEPRDLUTHPAGHAEN 367

QY 355 SLAYGRPHVCPGVSIARLEAEIANGVTIFRRPEMK-----LXETPFG 398
 DB 368 HVGFGHGAHYCLGATLAKQEGVALGALLRHFPPELSLAVAPALERTVPVG 418

RESULT 10
 AAB18654
 ID AAB18654 standard; Protein; 416 AA.

XX AAB18654;

DT 22-JUN-2001 (first entry)

DE Amino acid sequence of picromycin/methymycin cytochrome P450 hydroxylase.

XX Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
 KM antidiabetic; C12-hydroxylase; pick; desosamine biosynthesis;
 KM desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
 KM picromycin biosynthesis;
 KM picromycin/methymycin cytochrome P450 hydroxylase.


```

XX OS Streptomyces venezuelae.
XX XX US6117659-A.
XX PN 12-SEP-2000.
XX PD 27-MAY-1999; 99US-0320878.
XX PF 28-MAY-1998; 98US-0087080.
XX PR 22-SEP-1998; 98US-0100880.
XX PR 08-FEB-1999; 99US-0119139.
XX PR 20-MAY-1999; 99US-0134990.
XX PR 30-APR-1997; 97US-0846247.
XX PR 06-MAY-1998; 98US-0073538.
XX PR 28-AUG-1998; 98US-0141908.
XX XX
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX XX
XX PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX XX WPI; 2000-610844/58.
XX DR
XX XX New recombinant pick hydroxylase gene of Streptomyces venezuelae useful
XX PT for converting ketolids to antibiotics and as antibiotics and
XX PT intermediates in the synthesis of compounds with pharmaceutical value
XX XX
XX PS Disclosure; Columns 47-48; 117pp; English.
XX XX
XX CC The present sequence represents a picromycin/methymycin cytochrome P450
XX CC hydroxylase (pick). The nucleotide sequence encoding it is used in the
XX CC course of the invention. The specification describes a recombinant DNA
XX CC compound expressing recombinant polyketide synthase genes in host cells
XX CC for the production of narbonolide, narbonolide derivatives and
XX CC polyketides that are useful as antibiotics and as intermediates in the
XX CC synthesis of compounds with pharmaceutical value. The DNA compounds may
XX CC also encode a C12-hydroxylase (pick), desosamine biosynthesis and
XX CC desosaminyl transferase enzymes (useful for conversion of ketolids to
XX CC antibiotics), and the beta-glucosidase enzyme (involved in picromycin
XX CC biosynthesis). These compounds are also useful for increasing the
XX CC antibiotic activity of a compound relative to the unhydroxylated
XX CC compound. The recombinant host cells are useful as genetic systems that
XX CC allow rapid engineering of the narbonolide polyketide synthase. These
XX CC would be valuable for creating novel ketolide analogs for pharmaceutical
XX CC applications.
XX XX
XX SQ Sequence 416 AA;

```

```

OY 300 IIRIGTVPANODLEYCASIKKGMVFLIPSAIRDGTVSRPDPVVRDGTASLAVG 359
DB 289 PVESATYFRFPVVDLGTVPAGDTVLVVLADARTRPERPDPFRPDRRTAGHLAFG 348
OY 360 RGPVHCPSVSLARLAEIYAVGTIFRRFPDMLEKRP---VGYHAFPNIESLNV 411
DB 349 HGIFHCIGAPLARLEARIYVRLALERCPLDLVSPGELWVYPMIRGLKALPI 403

RESULT 11
AA67218
ID AA67218 standard; protein; 416 AA.
XX
XX AA67218;
AC 23-MAR-2000 (first entry)
XX
XX DT
XX DE Macrolide ring modifier pick hydroxylase amino acid sequence.
XX XX
XX KW Narbonolide polyketide synthase; PKS; hydroxylase; pick; ketolide;
XX XX antibiotic production; narbomycin; picromycin.
XX XX
XX OS Streptomyces venezuelae.
XX XX
XX PN WO9961599-A2.
XX PD 02-DEC-1999.
XX PF 27-MAY-1999; 99WO-US1814.
XX PR 28-MAY-1998; 98US-0087080.
XX PR 28-AUG-1998; 98US-0141908.
XX PR 22-SEP-1998; 98US-0100880.
XX PR 08-FEB-1999; 99US-0119139.
XX XX
XX PA (KOSA-) KOSAN BIOSCIENCES INC.
XX XX
XX PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L;
XX XX WPI; 2000-072618/05.
XX DR N-PSDB; AA256001, AA256003.
XX XX
XX PT New recombinant DNA encoding a domain of narbonolide polyketide
XX PT synthase, for production of ketolide antibiotics -
XX XX
XX PS Example 2; Page 40; 98pp; English.
XX XX
XX CC This is the Streptomyces venezuelae hydroxylase pick protein sequence.
XX CC The pick protein is involved in macrolide ring modification. The
XX CC invention relates to recombinant DNA containing a coding sequence for a
XX CC narbonolide polyketide synthase (PKS). Polyketides are compounds
XX CC synthesised from 2-carbon units through a series of condensations and
XX CC subsequent modifications. Modular PKSs are responsible for the
XX CC production of many antibiotics including picromycin. The narbonolide PKS
XX CC consists of a loading module, six extender modules, and two thioester
XX CC domains. Four proteins make up the narbonolide PKS (PICAI, PICAILI,
XX CC PICAIIT and PICAIIV). PICAI includes the loading module and extender
XX CC modules 1 and 2, PICAILI includes extender modules 3 and 4, PICAIIT
XX CC includes extender module 5 and PICAIIV includes extender module 6 and a
XX CC type II thioesterase domain. The second type II thioesterase domain is
XX CC found on the PICB protein. The nucleotide sequences encoding all of these
XX CC proteins can be isolated in recombinant form from the recombinant cosmid
XX CC PK05023-27 (see AA256001). Narbonolide is desosaminylated in S.
XX CC venezuelae to yield narbomycin, the desosaminyl transferase enzyme is
XX CC required for this conversion, and the desosamine biosynthetic genes are
XX CC also found in cosmid PK05023-27. The recombinant DNA of the invention is
XX CC used to express, in transformed cells, narbonolide (or its derivatives)
XX CC or other ketolides (particularly hybrids), which may then be converted
XX CC (e.g. by other enzymes recombinantly expressed in the same hosts) to
XX CC polyketide antibiotics or their intermediates. The antibiotics are useful
XX CC in human or veterinary medicine.
XX XX
XX SQ Sequence 416 AA;

```

Query Match	22.8%;	Score 484;	DB 21;	Length 416;
Best Local Similarity	30.6%;	Pred. No. 3.8e-38;		
Matches 127;	Conservative 81;	Mismatches 183;	Indels 24;	Gaps 8;

QY	5	QANQSEKPPAFDFEFPAGVDEDEFPPIERLEATPI--FYWDG3-RSMVLTTRHYADVAS	61
Db	5	QOGTTASPVLDDGALGODFPAADPYPTTARLRACGPAHRVTPESGDEWLVVGDRAAV	64
QY	62	FRDERFAVSREWESSAAYSASAIPELSDMKYGLFGLPEDHARVRLVNSFTSRAIDL	121
Db	65	LADPRFS---KDWKNS---TTPLEAEALNHNMLESDPPRHTRLRLKVAREFTMRAYEL	118
QY	122	LRAEIQRTVQQLDARSQGE--EDVYRDVAEGTIPMRAISALLKYPAECDEKFRFSGAT	179
Db	119	LRRVQVEITVGLVDAMLAAAPDGRADLMESLAMPPIITVISELIGVEPDRAAFRWTFAP	178
QY	180	ARALGVGLQVQDEBETKVLVASVTEGLALHDVLDERRNRNLENDVLTMLLQAEADGSR	239
Db	179	-----VFPDDPAQAQTAAKEMSGYISRL--IDSKKGGQGEIDLIALVTSDDGSR	228
QY	240	STKEIVALNGAIIAGDTDTTLYLLAFALNLTLSPEALELTKAEKGLRMAALDEVLR	299
Db	229	TSEELGMAHLLVAGHETTNYLNLANGVYALLSHPDQALARADMTLLDGAVENTLR	288
QY	300	ILRIQTVRFARQDLEYCGASIKKGEWYFLLLPSALRDGVTSRPDVEDVRDGTGASL	359
Db	289	PVESATYRFPEVPDLDTGTVIPADDTVLVLADAHRTPEREPDPRHFRIDRDTAGHLA	348
QY	360	RGRHVPCGVSLARLEAEIANGVITRRRPENMLKEP---VGCYHPAFNISLAV	411
Db	349	HGHFFCGAPLARLEARIYAVALLERCSDLLDVSPGLVWYNNMIGLALPI	403

RESULT 12
ABG71678
ID ABG71678 standard; protein; 416 AA.

AC ABG71678

DT 21-JAN-2003 (first entry)

S. venezuelae picromycin/methymycin cytochrome P450 hydroxylase, Pick.

25 narbonolide polyketide synthase; PKS; desosamine biosynthetic gene;
26 desosaminyl transferase gene; beta-glucosidase gene; anthiotoxic;
27 pick hydroxylase gene; c12 hydroxylase gene; narbonolide;
28 desosaminylated polyketide; narbonmycin biosynthesis; pick;
29 picromycin/methymycin cyclochrome P450 hydroxylase;
30 picromycin biosynthesis; enzyme.

Streptomyces venezuelae.

PN WO200297062-A2.

PD 05-DEC-2002.

PF 22-FEB-2002; 2002WO-US05642.

PR 22-FEB-2001; 2001US-0793708.

PA (KOSA-) KOSAN BIOSCIENCES INC.

PI Ashley G, Betlach MC, Betlach M, McDaniel R, Tang L,

WPI; 2003-041412/03.

PT Preparation of polyketides by recombinant DNA technology, useful as
PT antibiotics and as intermediates in the synthesis of pharmaceutical
PT compounds -

PS Disclosure; Page 45; 127pp; English.

Query Match	22.8%	Score 484	DB 24	Length 416	
Best Local Similarity	30.6%	Pred. No. 3.8e-38			
Matches 127	Conservative 81	Mismatches 183	Indels 24	Gaps 8	
Qy	5	QANSEKPAFDKFPAPGYAEDFPFAIERLREATPI - FYWDEG - RSWLTRYHDSAV	61		
Db	5	QQGTTASPVLDLIGALGDQDPAADPYPTYARLRAEGPAHRVTRTEGDEVMLVGYDRARAV	64		
Qy	62	FRDRFPVSKREWSSAEYSSAIPBELSDMKYGLFGIPREDNHAVRKLVNPSFISRALDL	121		
Db	65	LADRFs---KDWKNS---TTPLEBAALNNHNLSDPPRHTRKLVAREFTMRVEL	118		
Qy	122	LRAEIQGTVDQLDARSGOEE--FDVVRDYAEGIPMAISALILKVPACDECKEFPFGSAT	179		
Db	119	LRPVQGIIVGVNDMLAAPDGRADIMESLAWPIPIVIBELLGVPREDPAAPFVWDAF	178		
Qy	180	ARALVGLVPQVDEETKTLVASVTEGALLDHVLDERRRNPLENDVLTMLQEAQDSRL	239		
Db	179	-----VFPDDPAQAOTAMAEWSGYLSRL--IDSRRGDDGBDLISALVTSDEDSRL	228		
Qy	240	STKELVALVGAIIAGDITTYILAFVNLILRSPELDELKAPCGLMRNALDVLRFDN	299		
Db	229	TSEELIGMAHLLVAGHEITVNLIANGMVALLSHPDOLALRADMTLLDGAVEEMLYEG	288		
Qy	300	ILRIGTVRFARQDLEYCGASIKGEMVFLIPSLARDGTVFSRPDVEDVRDTSGLAYG	359		
Db	289	PVESATIRFPPEPVDLGGTVIPAGDYLVLVLAQDHRPEKFPDPHRRDIDRDGHLAFG	348		
Qy	360	RGPVVCQGVSLARLEAIIANGTIFRRRPEMKLQETP--VFGIYPAARNIESLNV	411		
Db	349	HGHFICIGAPLARLEARIAVRALIERCPDIALDVSPELWYPMPIRGKALPI	403		
RESULT 13					
ID	AAE34780	standard; Protein; 430 AA.			
XX	AAE34780				
AC	AAE34780				
XX	14-MAY-2003	(first entry)			
XX	DE	Streptomyces tubercidicus ema17 protein.			
XX	KW	P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;			
XX	KW	emamectin, insecticide; ema17 protein.			
OS	Streptomyces tubercidicus.				
XX	WO200292801-A2.				
XX	PD	21-NOV-2002.			
XX	PF	15-MAY-2002; 2002WO-EP05363.			
XX					

RESULT 13

ID AAE34780 standard; Protein; 430 AA.

AC AAE34780;

DT 14-MAY-2003 (first entry)

Streptomyces tubercidicus emal7 protein.

KW P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;

XX

XX

XX

[illegible][illegible]

XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
 PI Buckel TG;
 XX WPI; 2003-140280/13.
 DR N-PSDB; AAD53052.
 XX Novel polypeptide that exhibits enzymatic activity of P450
 PT monooxygenase and capable of regioselectively oxidizing alcohol of
 PT avermectin useful for making emamectin from avermectin -
 XX
 PS Claim 17; Page 147-148; 157pp; English.
 XX
 CC The present invention relates to novel proteins that exhibit an enzymatic
 CC activity of P450 monooxygenase and capable of regioselectively oxidizing
 CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
 CC The invention also relates to ferredoxins and ferredoxin reductases that
 CC are active with the P450 monooxygenases. Sequences of the invention are
 CC useful for producing 4-keto-avermectin from avermectin, which is useful
 CC for producing emamectin. Emamectin is useful as an insecticide. The
 CC present sequence is Streptomyces tubercidicus emat7 protein.
 XX
 SQ Sequence 430 AA;
 XX
 Query Match 22.6%; Score 480.5; DB 24; Length 430;
 Best Local Similarity 29.1%; Pred. No. 8,8e-38;
 Matches 130; Conservative 89; Mismatches 161; Indels 67; Gaps 12;
 QY 14 AEDFKPFA-----PGVADEPPAIEERLEATPIF--YWDGSGWVLTTRYH 56
 DB 3 AISSSPFAHVGHKPGEPVMDPALITDPFGGALREGGYPVLRGFMDSPLYWLTTRYE 62
 QY 57 DVSAVFRDRFAVSRREWESSAAYSALPELSDMKYGLFGLPP-----101
 DB 63 EVROVLRDQRFNN---PAASPGHSDIESPTARLLDMGMGEPHFRPYLMSILNNDAP 118
 QY 102 DHARVKLVNPSPTSAIDLLRAEIQRTVDOLLDAISGQEE--FVYVADYAEIGPMRAI 158
 DB 119 DHTLRRLVSRVATARKITDLRPRVQLADELRLRPEHAEQGVNVLINFAIPLETVI 178
 QY 159 SALLKVPACDEKFRFGSATRALQVGLVQVDEETKTLVASVTEGLALLHDVDERRR 218
 DB 179 CELVGPEDRPRQMRKMG-----DLVSDPELSTSPFMIHE--IHLEIFE-RR 226
 QY 219 NPLENDVLTMLIA--EADGSRISTEVALVGAIIAGTDTTYLITAFVNLILRSPAL 277
 DB 227 GALTDLLELIRTHDDGSRISDVEMVTVLVLAGHETTHALINGSTAAALLTRPDOL 286
 QY 278 ELVKAEPGLMRNALDVAFEDNLRIGTVRFARODLEYGASIKKEMFLILPSALRGG 337
 DB 287 RLVKQDPPELLPRVNHLLRMCQPGVQOTQRYASEDEVLIGTPRKGDVAQLILVSNFDP 346
 QY 338 TVFSRDPDVVRDGA---SLAYGRPHVCPGVSLARLEAIVAGTIFRRFPPEM-----389
 DB 347 RHYTAERDLTRHPRGHAENHYVFGHNGYCGAILLAGEGVAVAGKLTHTYBELSLAV 406
 QY 390 ---KKEETPVFGVHPRFNIESINVL 413
 DB 407 APDELEETPV----PGSWRLDSLPRVL 429

RESULT 14

AA77196

AA77196 standard; Protein; 415 AA.

AA77196; (first entry)

S. venezuelae macrolide biosynthetic enzyme pikC, SEQ ID NO:39.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
 KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
 KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
 KW chronic obstructive pulmonary disease; respiratory inflammation;
 KW hypercholesterolemia; crop protection agent.
 XX
 OS Streptomyces venezuelae ATCC15439.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 103
 FT /note= "Encoded by CCG"

WO200000620-A2.

06-JAN-2000.

25-JUN-1999; 99WO-US14398.

26-JUN-1998; 98US-0105537.

(MINU) UNIV MINNESOTA.

Sherman DH, Liu H, Xue Y, Zhao L;

WPI; 2000-160679/14.

N-PSDB; AA287301.

Desosamine and macrolide biosynthetic gene clusters, useful for, e.g.

synthesis of methymycin and pikromycin -

Claim 19; Page 429-430; 438pp; English.

XX The invention relates to an isolated and purified nucleic acid segment
 CC comprising a desosamine biosynthetic gene cluster, a fragment or its
 CC biologically active variant, where the nucleic acid sequence is not
 CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
 CC streptomycetes antibiotics. The invention also relates to a macrolide
 CC biosynthetic gene cluster, or fragments thereof. The macrolide
 CC biosynthetic gene cluster encodes proteins which synthesize methymycin,
 CC pikromycin, neomethymycin, narbomycin or a combination of these
 CC compounds. Recombinant or augmented cells comprising the desosamine
 CC and/or macrolide biosynthetic gene clusters are useful for the
 CC production of biologically active macrolides. The macrolide biosynthetic
 CC proteins are useful for synthesis of methymycin, pikromycin,
 CC neomethymycin and narbomycin. The alternative termination of polyketide
 CC synthesis may be useful to prepare novel antibiotics and
 CC polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
 CC recombinant host cells are useful as biopolymers, e.g., in packaging or
 CC biomedical applications, to engineer PHA monomer synthases or to prepare
 CC biologically active agents, such as chemotherapeutics.
 CC immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
 CC disease as well as other diseases involving respiratory inflammation,
 CC cholesterol-lowering agents or macrolide-based antibiotics which are
 CC active against a variety of organisms, e.g., bacteria, including
 CC multi-drug resistant pneumococci and other respiratory pathogens, as well
 CC as viral parasitic pathogens, or as crop protection agents (e.g.,
 CC fungicides or insecticides) via expression of polyketides in plants.
 CC Sequences AA77190-77197 represent macrolide biosynthetic enzymes from
 CC Streptomyces venezuelae ATCC 15439, which are encoded by sequences
 CC AA287295-287302.
 XX
 SQ Sequence 415 AA;

Query Match 22.4%; Score 476.5; DB 21; Length 415;

Best Local Similarity 30.5%; Pred. No. 2e-37; Indels 25; Gaps 9;

Matches 128; Conservative 81; Mismatches 185;

1 MTOGANOSETKAPDFKPAFGVADPPAIEERLEATPI--FWWDG-BSWVLTTRYAD 57

1 MRRTOGTTASPPVLDLALGDPFAADYPTTARLARAGPARRVTPEDGDMVLVGYGR 60

58 VSAVFRDRFAVSRREWESSAAYSALPELSDMKYGLFGLPPEDHARVKLVNDSFTSR 117

```

Db      61 ARAVLADPPRES---KQWRNS---TTPLEAEALNNHMLSDPPRHTR-RKLVAEFTMR 113
Qy      118 AITLLAEIORTYDQLLDASGOEE--FDVVRDYAEGIPMRALSALKVPAECDEKRRF 175
      114 RVLLLRPRVEIYDGLVDALAAPDGRADLMESLAMPPTIVISELGVPEPPRAARVW 173
Qy      176 GSATARALAGVLPOVDEETKTIVASVTEGLALHLDVDERRRNPLENDVLTMLQEA 235
      174 TDAF-----VPPDDPAQQTMAEMSGVLSRL---IDSKRGQDGDGLSALVRSDED 223
Qy      236 GSRLSTKEVLVALGAIIAAGTDTTIVLIAFAVNLRLSPALBLVKAEPGLMRNALDEV 295
      224 GSRLTSEBELGMHIIILVAGHETTIVLIANGMYALLSHPPQLAALRADMTILDGAVEML 283
Qy      296 RPNILIRIGTRPARODLEYCGASIKKGEMVFLIPSLADGTVEFSRSDVEVDVRRDTGAS 355
      284 RYGPVESATYRRPVEPVDDGTVPAGDTVLVLADAHRTPEPFPDPHFDLRDPAH 343
Db      356 IAYGRGPHVCPGVSILAEIAVGTIFRRFPPEMKLKEP---VFGYHAPFNNIESIN 411
      344 LAFGHGHCICGAPLARLEARIANRALREKCPDLADVSGELVWYFNPWIRGLKALPI 402

RESULT 15
ID      AAE34725 standard; Protein; 430 AA.
AC      AAE34725;
XX      14-MAY-2003 (first entry)
DE      Streptomyces lydicus ema4 protein.
XX      DE
KM      P450 monooxygenase; avermectin; ferrodoxin; ferrodoxin reductase; enzyme;
KM      emamectin; insecticide; ema4 protein.
XX      KM
OS      Streptomyces lydicus.
XX      OS
PN      MO200292801-A2.
XX      PN
PD      21-NOV-2002.
XX      PD
PF      15-MAY-2002; 2002MO-EP05363.
XX      PF
PR      16-MAY-2001; 2001US-291149P.
XX      PR
PA      (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX      PA
PI      Molnar I, Ligon JM, Zirkle RE, Hammer PE, Hill DS, Pachlatko JP;
PI      Buckel TG;
XX      PI
DR      WPI; 2003-140280/13.
XX      DR
DR      N-PSDB; AAD53020.
XX      DR
PT      Novel polypeptide that exhibits enzymatic activity of P450
PT      monooxygenase and capable of regioselectively oxidizing alcohol of
PT      avermectin useful for making emamectin from avermectin -
XX      PT
PS      Claim 17; Page 109-110; 157pp; English.
XX      PS
XX      CC The present invention relates to novel proteins that exhibit an enzymatic
XX      CC activity of P450 monooxygenase and capable of regioselectively oxidizing
XX      CC the 4-carbinol group at position 4 of avermectin to 4-keto-avermectin.
XX      CC The invention also relates to ferrodoxins and ferrodoxin reductases that
XX      CC are active with the P450 monooxygenases. Sequences of the invention are
XX      CC useful for producing 4-keto-avermectin from avermectin, which is useful
XX      CC for producing emamectin. Emamectin is useful as an insecticide. The
XX      CC present sequence is Streptomyces lydicus ema4 protein.
XX      CC
SQ      Sequence 430 AA;

```

Query Match 22.4%; Score 475.5; DB 24; Length 430;
 Best Local Similarity 30.9%; Pred. No. 2.7e-37;

```

Matches 127; Conservative 81; Mismatches 152; Indels 51; Gaps 13;
Qy      22 PGVAEDPFAIRLEIRATPIF---YWDGRSWLTYHDVSAVFRERPAVSREWESSA 78
      25 PALIGDFGCGALRQGVVAGREFNDSPWFVIRFEVREYLRDPRF---RNNVSA 81
Qy      79 EVSSAIPELSDMKK-YGLGLPPE-----DHARVRKLVNPSFTSRAIDLL 122
      82 --PGAAPEDTPLSRLLMDMMGFPEHLRVLLGSIINNDADPHTRRLVSRAPFAFKITDL 139
Qy      123 RAEIQTVDQLLDASGOEE---FDVVRDYAEGIPMRALSALKVPAECDEKRRRGSAT 179
      140 RPRVTOIADBELARLEBEHAEDGVVDLIQFAFPLPTIVICELVGIPEDRPQWRWGADL 199
Qy      180 ARAVGVLPOVDEETKTIVASVTEGLALHLDVDERRRNPLENDVLTMLQEA-BA 238
      200 ----VSLDP--DRMSRSFPAMIDH---IHELIARRR-ALTDDLSELIRTHDDGSR 247
Qy      239 LSTKELVALGAIIAAGTDTTIVLIAFAVNLRLSPALBLVKAEPGLMRNALDEVLPD 298
      248 LSDVENVTVLTVLVAHETTALHNGTAAALLTHPDQLRLKDDPALLPRAVHELMRWC 307
Qy      299 NILRIGTRPARODLEYCGASIKKGEMVFLIPSLADGTVEFSRSDVEVDVRRDTGA---- 354
      308 GPVHMTQLRYAAEDVELAGVRIKRTGDAVOLIVSANRDRPHYDLPDRDLTRHPAGHAEN 367
Qy      355 SLAYGRGPHVCPGVSILAEIAVGTIFRRFPPEMK-----LKETPVFG 398
      368 HVGFGHAHYCIGATLAKOGEVALGALRHPEBELSLAFAPEALENTPPFG 418

```

Search completed: October 2, 2003, 17:33:53
 Job time : 30.237 secs